УДК 004.8:78.01

ЗАСТОСУВАННЯ МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ МУЗИЧНИХ ЖАНРІВ І ПІДЖАНРІВ

Олег Вишневський

Миколаївський національний аграрний університет, м. Миколаїв, Україна

Науковий керівник: О. Є. Богатєнкова,
асистент кафедри економічної кібернетики,
комп'ютерних наук та інформаційних технологій
Миколаївський національний аграрний університет

Стрімкий розвиток цифрових технологій і збільшення обсягів музичного контенту зумовлюють потребу в автоматизованих методах його аналізу та систематизації. Сучасні музичні платформи містять значну кількість аудіозаписів, які необхідно впорядковувати за жанрами, стилями та іншими характеристиками. Традиційна класифікація музичних творів переважно передбачає участь людини, що потребує часу та може залежати від суб'єктивного сприйняття. У зв'язку з цим особливого значення набувають методи машинного навчання, здатні автоматично визначати характерні ознаки музичних композицій на основі аналізу аудіосигналів.

Одним із перспективних напрямів розвитку таких систем є застосування глибокого навчання. На відміну від традиційних підходів, які часто потребують попереднього ручного визначення характеристик звукового сигналу, нейронні мережі можуть самостійно виявляти складні закономірності у вхідних даних. Це створює передумови для розроблення моделей, здатних розпізнавати не лише загальні музичні жанри, а й окремі піджанри, що мають подібні акустичні характеристики. Метою дослідження є теоретичне вивчення можливостей застосування методів глибокого навчання для класифікації музичних жанрів і піджанрів на основі аналізу аудіосигналів, визначення основних етапів обробки звукових даних та обґрунтування доцільності використання нейронних мереж у цьому процесі.

Музичний жанр є категорією, яка характеризує твір за сукупністю властивостей, зокрема інструментальним складом, ритмічною організацією, гармонічною структурою та особливостями звучання. Як зазначають Дж. Танакіс і П. Кук, автоматична класифікація музики може ґрунтуватися на аналізі темброві, ритмічних і тональних характеристик аудіосигналів [1]. Такий підхід став важливим підґрунтям для подальшого розвитку систем пошуку та автоматизованої класифікації музичного контенту.

Класифікація музичних піджанрів є складнішим завданням, оскільки відмінності між ними не завжди мають чітко виражений характер. Наприклад, композиції в межах рок- і метал-музики можуть використовувати подібні інструменти, темпи, ритмічні структури та прийоми виконання. Водночас на їхню стилістичну належність впливають характер гітарного звучання, особливості вокалу, ритмічні рисунки, структура композиції та співвідношення різних звукових компонентів. Тому для надійного розпізнавання піджанрів необхідно враховувати сукупність ознак, а не спиратися на одну окрему характеристику.

Основою автоматизованої класифікації є цифровий аудіосигнал, який представляє звукові коливання у вигляді послідовності числових значень. Безпосереднє використання повного аудіозапису як вхідних даних може вимагати значних обчислювальних ресурсів, особливо якщо композиції мають велику тривалість. Тому перед передаванням даних до моделі зазвичай здійснюють попередню обробку: декодування аудіофайлу, приведення сигналу до узгодженої частоти дискретизації, за потреби перетворення стереозапису на монофонічний сигнал і поділ композиції на фрагменти однакової або контрольованої тривалості. Це дає змогу уніфікувати вхідні дані та спростити їх подальший аналіз.

Важливим етапом є представлення звукової інформації у формі, придатній для обробки нейронною мережею. Одним із поширених засобів такого представлення є спектрограма, яка відображає розподіл енергії сигналу за частотами в часі. На відміну від звичайного часового графіка

аудіосигналу, спектрограма дає змогу досліджувати частотну структуру звуку та зміни його спектрального складу протягом композиції.

Для музичної класифікації особливий інтерес становить мел-спектрограма, побудована з використанням мел-шкали, яка наближено враховує особливості сприйняття частот людським слухом. Таке представлення може допомагати виявляти відмінності між звуковими текстурами, характером інструментального звучання та іншими акустичними властивостями. Мел-спектрограми часто використовують як вхідні дані для згорткових нейронних мереж, оскільки їхня двовимірна структура дає змогу аналізувати локальні закономірності одночасно в часовому та частотному вимірах [2].

Іншим поширеним способом опису аудіосигналу є мел-частотні кепстральні коефіцієнти MFCC. Вони формують компактне представлення спектральних характеристик звуку та традиційно застосовуються в задачах розпізнавання мовлення й аналізу музики. Використання MFCC дає змогу зменшити розмірність ознак і зосередитися на певних властивостях тембрової структури. Вибір між MFCC, мел-спектрограмами або іншими представленнями залежить від особливостей навчальної вибірки, архітектури моделі та поставленого завдання.

Для обробки таких даних перспективним інструментом є згорткові нейронні мережі Convolutional Neural Networks. Їхня робота ґрунтується на застосуванні згорткових операцій, які дають змогу виявляти локальні закономірності у вхідних даних. На початкових етапах навчання мережа може формувати детектори простих спектральних структур, тоді як глибші шари поєднують отримані ознаки у складніші представлення. У контексті музики це створює можливість аналізувати комбінації акустичних характеристик, які можуть бути пов’язані з певними жанровими особливостями. Дослідження К. Чоя, Дж. Фазекаса та М. Сендлера розглядає особливості роботи глибоких згорткових мереж під час класифікації музики на основі спектрограм [2].

Процес навчання моделі передбачає використання набору аудіозаписів, для кожного з яких відомий відповідний жанр або піджанр. Під час навчання нейронна мережа отримує вхідні дані, формує прогноз і порівнює його з правильною міткою. На основі визначеної помилки коригуються параметри моделі, що поступово дає змогу покращувати відповідність між акустичними ознаками та класами музичних творів. Для багатокласової класифікації вихідний шар може формувати оцінки належності композиції до кожного з визначених класів, після чого обирається клас із найвищою оцінкою.

Водночас якість класифікації залежить не лише від архітектури нейронної мережі, а й від характеристик навчальних даних. Вибірка повинна достатньо повно представляти кожен із досліджуваних піджанрів. Якщо певні класи містять значно більше аудіозаписів, ніж інші, модель може надавати перевагу поширенішим категоріям. Додатковою проблемою є правильність жанрового маркування, оскільки одна композиція може поєднувати ознаки кількох стилів, а її класифікація іноді залежить від обраних критеріїв.

Особливу увагу необхідно приділяти поділу даних на навчальну, валідаційну та тестову вибірки. Навчальна вибірка використовується для налаштування параметрів моделі, валідаційна допомагає добирати архітектуру й гіперпараметри, а тестова призначена для підсумкового оцінювання якості класифікації. Якщо фрагменти однієї композиції потрапляють одночасно до різних вибірок, модель може запам’ятати специфічні характеристики запису замість того, щоб навчитися розпізнавати загальні ознаки піджанру. Аналогічна проблема виникає, коли в навчальних і тестових даних представлені композиції тих самих виконавців. Тому для об’єктивнішої перевірки здатності моделі узагальнювати закономірності доцільно, за можливості, розділяти дані з урахуванням виконавців.

Для оцінювання ефективності класифікації використовують низку кількісних показників. Accuracy характеризує частку правильно класифікованих прикладів серед усіх тестових даних. Precision відображає точність віднесення композицій до певного класу, Recall показує, яку частину всіх композицій цього класу вдалося правильно виявити, а F1-score узагальнює Precision і Recall за допомогою їхнього гармонійного середнього. У багатокласових задачах ці показники можна обчислювати окремо для кожного класу та узагальнювати різними способами. Додатковим засобом аналізу є матриця помилок, яка дає змогу визначити, між якими саме піджанрами найчастіше виникає

плутанина. Застосування кількох показників забезпечує повніше уявлення про сильні та слабкі сторони моделі, особливо за нерівномірного представлення класів.

Таким чином, застосування глибокого навчання для класифікації музики передбачає поєднання методів цифрової обробки аудіосигналів, ефективного представлення акустичних ознак і навчання нейронних мереж. Для завдань розпізнавання рок- і метал-піджанрів особливого значення набуває здатність моделі враховувати складні комбінації тембрових, ритмічних і спектральних характеристик. Водночас отримання надійних результатів потребує якісно сформованої навчальної вибірки, коректного розподілу даних та комплексного оцінювання моделі.

Методи глибокого навчання є перспективним інструментом автоматизованої класифікації музичних жанрів і піджанрів. Використання мел-спектрограм, MFCC та згорткових нейронних мереж створює можливості для виявлення характерних закономірностей в аудіосигналах без необхідності вручну визначати всі ознаки музичного твору. Водночас ефективність класифікації залежить від репрезентативності даних, якості жанрового маркування, обраної архітектури та способу перевірки результатів. Подальше дослідження цих підходів може стати теоретичною основою для розроблення систем автоматичного розпізнавання піджанрів рок- і метал-музики.

Список використаних джерел:

1. Tzanetakis G., Cook P. Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing. 2002. Vol. 10, No. 5. P. 293-302.

URL: https://doi.org/10.1109/TSA.2002.800560.

2. Choi K., Fazekas G., Sandler M. Explaining Deep Convolutional Neural Networks on Music Classification. arXiv preprint arXiv:1607.02444. 2016.

URL:https://www.researchgate.net/publication/305119391_Explaining_Deep_Convolutional_Neural_Networks_on_Music_Classification

3. McFee B., Raffel C., Liang D., Ellis D. P. W., McVicar M., Battenberg E., Nieto O. librosa: Audio and Music Signal Analysis in Python. Proceedings of the 14th Python in Science Conference. 2015 P. 18-25. URL: https://doi.org/10.25080/Majora-7b98e3ed-003.

УДК 004.85:004.89

ЗАЛЕЖНІСТЬ ТОЧНОСТІ ПРОГНОЗУВАННЯ ТЕОРЕТИЧНОГО РЕЗУЛЬТАТУ ПОЗИЦІЙ ГРИ «ЧОТИРИ В РЯД» ВІД ОБСЯГУ НАВЧАЛЬНОЇ ВИБІРКИ

Людмила Венгура, Олександр Пархоменко

Миколаївський національний аграрний університет, м. Миколаїв, Україна

У клієнт-серверній інтелектуальній системі гри «Чотири в ряд» серверна частина отримує від клієнта поточну позицію й повертає хід або її оцінку. Теоретичний результат позиції точно визначають розв'язувачі [1–4], однак без бази дебютів на це йде орієнтовно 0,7–1 с на позицію на одному ядрі процесора (власні вимірювання), що ускладнює масові запити. Модель машинного навчання, навчена на розмічених розв'язувачем позиціях, може слугувати швидкою оцінною функцією, але обсяг розмічених даних визначає і вартість їх підготовки, і якість оцінки. Тому доцільно кількісно оцінити, як точність залежить від обсягу навчальної вибірки.

В архітектурі такого типу клієнт (веб-інтерфейс) надсилає на сервер стан дошки, а сервер поєднує два компоненти: розв'язувач, що дає точний результат і використовується для розмітки даних, та класифікатор, що за мілісекунди повертає наближену оцінку. Гра розв'язана: у 1988 р. Allis показав, що перший гравець виграє, починаючи з центрального стовпця [1]; Tromp згодом визначив результати для різних розмірів дошки [3] і оприлюднив набір усіх позицій восьмого напівходу з теоретичними оцінками [2], який використовують у задачах класифікації. Саме така задача – передбачення теоретичного результату за станом дошки – відповідає ролі швидкого оцінювача на сервері.