

О. М. Ткаченко, к.т.н., доц., О. Ф. Грійо Тукало, асп.

ІДЕНТИФІКАЦІЯ АУДІОЗАПІСУ ЗА КОРОТКИМ ФРАГМЕНТОМ НА ОСНОВІ ПРИВЕДЕНОЇ ВЛАСНОЇ ВІДСТАНІ

В сучасних системах обробки аудіоінформації виникає необхідність автоматичного швидкого пошуку музичних творів на основі аудіоконтенту у віддалених базах даних (БД) великого розміру. Важливо, враховуючи великі обсяги аудіоінформації в БД, щоб тривалість фрагменту була якомога меншою, оскільки це дозволить: 1. збільшити швидкість пошуку; 2. зменшити час завантаження та мережевий трафік. Разом з тим зменшення тривалості фрагменту може зумовити зростання ймовірності помилки під час ідентифікації музичного твору. Тому метою роботи є теоретичне обґрунтування можливості ідентифікації музичного твору за коротким фрагментом, що дозволяє зменшити складність обчислень в процесі автоматичної ідентифікації музичного твору.

Схема визначення подібності музичних творів на основі контенту базується на використанні аудіофайлу для побудови моделі аудіосигналу. Зазвичай відліки аудіо сигналу ділять на невеликі фрагменти (фрейми) тривалістю 10 – 30 мс. Для кожного фрейму виконується спектральний аналіз, на основі якого (тим чи іншим чином) обчислюється значення вектора параметрів (параметризація). Для опису аудіосигналу як параметри було обрано мел-частотні кепстральні коефіцієнти (MFCC), що є компактним представленням спектральної обвідної. Критерієм порівняння музичних творів є евклідова відстань D .

Враховуючи періодичний характер музичних творів, що полягає в наявності схожих за текстом та характером мелодії фрагментів, в процесі формування шаблонів музичних творів в БД було застосовано кластерний аналіз, що дозволив скоротити кількість параметрів для опису шаблону музичного твору та відповідно зменшити обсяги пам'яті для їх зберігання.

Послідовності векторів параметрів MFCC, що описують музичні твори, було кластеризовано, використовуючи вдосконалений метод кластеризації k -середніх, який, на відміну від класичного, дозволяє отримати розв'язок, наближений до глобального мінімуму спотворення шляхом послідовного запуску k -середніх для центроїдів 1, 2, ..., k . Кожен шаблон БД було представлено 1000 кластерів MFCC. Очевидно, що відстань між файлами параметрів, що описують один і той же музичний твір до кластеризації і після (відстань до власного шаблону), буде додатною і рівною величині сумарної похибки кластеризації E^2 .

Під час кластеризації обчислюється однакова кількість кластерів для усіх шаблонів, однак музичні твори мають різну тривалість (кількість фреймів). Це призводить до того, що для творів, тривалість яких більша, початкова похибка (між файлами того ж твору до і після кластеризації) теж буде більшою, оскільки в цьому випадку на кожен кластер буде припадати більше векторів параметрів. Позбутися цього можна за рахунок ділення відстані до власного шаблону БД, на кількість

фреймів пісні n , що по суті є математичним очікуванням (МО) сумарної похибки кластеризації E^2 .

Цю величину було названо приведеною власною відстанню ($D_{ПВ}$) музичного твору. $D_{ПВ}$ не залежить від кількості фреймів (тривалості запису). Таким чином, $D_{ПВ}$ можна використовувати як критерій прийняття рішення в процесі ідентифікації музичного твору як на основі запису в цілому, так і на основі його фрагменту.

Отже, в роботі теоретично обґрунтовано можливість ідентифікації музичного твору за коротким фрагментом на основі приведеної власної відстані (МО похибки кластеризації), значення якої не залежить від кількості фреймів (тривалості запису). Визначено мінімальну тривалість фрагменту (5с), що дозволяє зменшити складність обчислень в десятки разів в процесі автоматичної ідентифікації музичного твору. Запропоновано критерій порівняння фрагменту музичного твору з шаблонами БД. Проведено експериментальні дослідження, що підтвердили коректність теоретичних положень.