Ідентифікація фрагмента музичного твору на основі приведеної власної відстані

Ткаченко О. М., Грійо Тукало О. Ф.

Вінницький національний технічний університет
Вінниця, Україна
alextk1960@gmail.com, xxmargoxx@gmail.com

Анотація

Робота належить до області інформаційних технологій, зокрема автоматичної ідентифікації музичного твору на основі аудіоконтенту. Теоретично обґрунтовано можливість ідентифікації музичного твору за його фрагментом. Під час формування бази даних шаблонів музичних творів застосовано кластерний аналіз, що дозволило зменшити обсяги пам'яті для їх зберігання. Запропоновано критерій порівняння фрагменту музичного твору з шаблонами бази даних. Визначено мінімальну тривалість фрагменту музичного твору, необхідну для його ідентифікації, що дозволяє суттєво зменшити складність обчислень в процесі пошуку музичного твору в базі даних. Експериментальні результати підтвердили коректність теоретичних положень.

1. Вступ

Дана робота присвячена вирішенню задачі автоматичної ідентифікації музичного твору за фрагментом аудіозапису. Важливо, 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. Таким чином, метою роботи є теоретичне обґрунтування можливості ідентифікації музичного твору за його фрагментом та мінімальної тривалості фрагменту, що дозволяє зменшити складність обчислень в процесі автоматичної ідентифікації музичного твору.

2. Вибір методу порівняння невідомого музичного твору з шаблонами БД

В роботі як параметри обрані мел-частотні кепстральні коефіцієнти (MFCC – Mel Frequency Cepstral Coefficients) [1-3]. Обравши MFCC як параметри, ми отримуємо опис музичного твору у вигляді файлу з параметрами MFCC. Параметризація (за допомогою MFCC) дозволяє в десятки разів зменшити кількість інформації, необхідної для опису музичного твору.

Загальну схему ідентифікації музичного твору наведено на рисунку 1. Найпростішим і очевидним підходом для визначення близькості між наборами параметрів MFCC невідомого музичного твору та еталонів БД є порівняння MFCC на основі Евклідової метрики, точніше квадрату Евклідової відстані $D_{Eu}{}^2$ (щоб надати велику вагу більш віддаленим об'єктам). Відповідно відстань між файлами параметрів невідомого твору та шаблону БД можна знайти за формулою:

$$D(\mathbf{X},\tilde{\mathbf{Y}}) = \sum_{j=1}^{n} D_{Eu}{}^2(\mathbf{x}_j,\tilde{\mathbf{y}}_j) = \sum_{j=1}^{n}\sum_{i=1}^{d}\left(x_{ji}-\tilde{y}_{ji}\right)^2. \qquad (2.1)$$

В ідеальному випадку при такому підході відстань між файлами параметрів MFCC одного і того ж музичного твору буде рівна нулю $D(\mathbf{X},\tilde{\mathbf{X}})=0$, для різних творів – відмінною від нуля $D(\mathbf{X},\tilde{\mathbf{Y}})>0$. Проте навіть аудіозаписи одного і того ж музичного твору можуть відрізнятися, наприклад: на початку запису може йти тиша, мелодія іншого музичного твору тощо; записи можуть мати різний темп, тривалість.

Це означає, що у разі зсуву фреймів в часі відстань до власного шаблону $D(\mathbf{X},\tilde{\mathbf{X}})\neq 0$, тобто умова чіткого розрізнення не виконується, отже, безпосереднє порівняння файлів параметрів за Евклідовою відстанню не підходить для задачі ідентифікації власного шаблону. В цьому випадку придатним є алгоритм динамічної трансформації шкали часу (DTW) [4]. Однак використання DTW призведе до зростання кількості операцій порівняння, що є неприйнятним.

Очевидно, що в більшості випадків музичний твір характеризується певною періодичністю, що полягає в наявності ідентичних або дуже схожих за текстом та характером мелодії фрагментів. Відповідно можна говорити про надлишковість даних, якими описується музичний твір, і можливість скоротити кількість параметрів для його опису. З огляду на це доцільним є застосування методів кластерного аналізу. Використання кластеризації для формування еталонів, що містяться в БД, дозволить зменшити обсяги пам'яті, необхідні для їх зберігання. Задачу кластеризації можна сформулювати так: заданий набір з n векторів, кожен з яких має розмірність d, необхідно розбити на підмножини відповідно до заданого критерію оптимізації. Як правило, таким критерієм є мінімізація спотворення $e_i^2 \to \min$. В більшості прикладних реалізацій для оцінювання спотворення використовують суму середньоквадратичних Евклідових відстаней між центром кластеру (центроїдом) $\mathbf{c}_i$ і векторами параметрів, які до нього належать $\mathbf{X}_i=\{\mathbf{x}\}, \mathbf{X}_i\subset\mathbf{X}$ [5, 6], тобто:

$$e_i^2 = \{\mathbf{c}_i : \sum_{j=1}^{N_i} D_{Eu}{}^2(\mathbf{x}_j,\mathbf{c}_i)\,|\,\mathbf{x}\in\mathbf{X}_i \le \sum_{j=1}^{N_i} D_{Eu}{}^2(\mathbf{x}_j,\mathbf{c})\,|\,\mathbf{x}\in\mathbf{X}_i\},$$

$$\mathbf{X}_i\subset\mathbf{X}, \forall\mathbf{c}\in\mathbf{X}\setminus\mathbf{X}_i, e_i^2\to\min.$$

де N_i – кількість векторів, що належать центроїду $\mathbf{c}_i$.

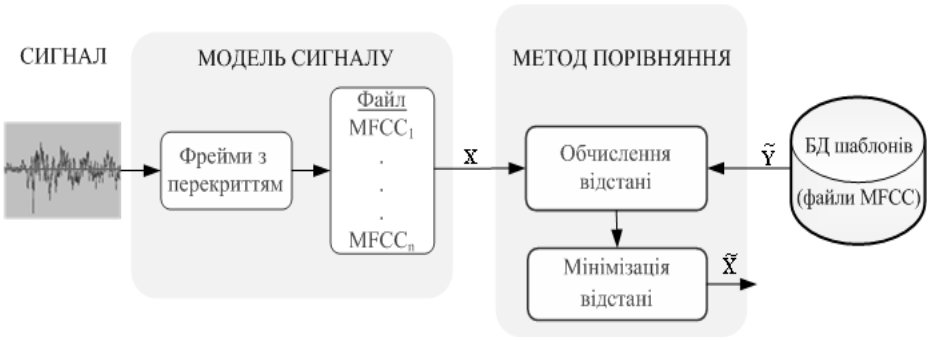

Рисунок 1. Загальна схема ідентифікації музичного твору

Таким чином, кожен шаблон було представлено 1000 кластерів (в середньому на кластер припадає близько 20 векторів).

Основні етапи порівняння файлів параметрів невідомого музичного твору з певним шаблоном БД:

1. Пошук мінімальної евклідової відстані $D\min^2$ між поточним вектором параметрів $\mathbf{x} = (x_1, x_2, ..., x_d)$ з множини параметрів $\mathbf{X} = \{\mathbf{x}\}, |\mathbf{X}| = n$ музичного твору, який треба ідентифікувати, та множиною векторів-кластерів $\tilde{\mathbf{Y}} = \{\tilde{\mathbf{y}}\}, |\tilde{\mathbf{Y}}| = m$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_d)$ шаблону:

$$D\min^2 = \underbrace{\min}_{j}(D_{Eu}{}^2(\mathbf{x}, \tilde{\mathbf{y}}_j)),\ j = \overline{1,m}. \qquad (2.2)$$

2. Обчислення оцінки відстані в цілому до шаблону $D\min^2$:

$$D(\mathbf{X}, \tilde{\mathbf{Y}}) = \sum_{l=1}^{n} D\min^2 = \sum_{l=1}^{n} \underbrace{\min}_{j}(\sum_{i=1}^{d}(x_i - \tilde{y}_{ji})^2),\ j = \overline{1,m}. \qquad (2.3)$$

Між кожним кластером та векторами, які належать до нього, є похибка e_i^2 (сума квадратів відстаней між ними). Звідси випливає, що відстань між файлами параметрів, що описують один і той же музичний твір до кластеризації і після (навіть якщо аудіозаписи були ідентичними), буде додатною і рівною величині сумарної похибки кластеризації E^2. Таким чином, відстань до власного шаблону дорівнюватиме:

$$D(\mathbf{X}, \tilde{\mathbf{X}}) = \sum_{l=1}^{n} D\min^2 = E^2 = \sum e_i^2, E^2 \to \min \qquad (2.4)$$

3. Оцінювання похибки за приведеною власною відстанню

Очевидно, що оскільки твори мають різну тривалість, кожний твір характеризується власною кількістю фреймів, представлених параметрами MFCC. Під час кластеризації обчислюється однакова кількість кластерів для усіх шаблонів. 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, визначеної в формулі (2.4), на кількість фреймів музичного твору n. Назвемо цю величину приведеною власною відстанню ($D_{ПВ}$) музичного твору:

$$D(\mathbf{X}, \tilde{\mathbf{X}}) = D_{ПВ} = \frac{\sum_{l=1}^{n} D\min^2}{n} = \frac{\sum_{l=1}^{n} \min(\sum_{i=1}^{d}(x_i - \tilde{y}_{ji})^2)}{n} = \frac{E^2}{n},\ j = \overline{1,m}, \qquad (3.1)$$

Таким чином, $D_{ПВ}$ – по суті є математичним очікуванням (МО) похибки кластеризації:

$$D_{ПВ} = M_E = \frac{E^2}{n}. \qquad (3.2)$$

Як можна побачити з формули (3.1) характеристика $D_{ПВ}$ не залежить від кількості фреймів (тривалості запису). Таким чином, її можна використовувати як критерій прийняття рішення як для аудіозапису в цілому, так і для його окремого фрагменту. Проте це твердження буде справедливим, тільки якщо для різних фрагментів ця характеристика буде змінюватися незначно, тобто за умови стаціонарності процесу.

4. Обґрунтування можливості ідентифікації музичного твору за його фрагментом

Змінення значень похибки, що виникає при порівнянні певного запису музичного твору з шаблоном, є випадковою функцією (процесом), що протікає в часі E(t). Значення похибки кожного музичного твору є окремими реалізаціями випадкової функції E(t). Виходячи з самого принципу формування кластерів можна очікувати, що процес змінення значень похибки кластеризації в часі буде носити стаціонарний характер. Проте це припущення потребує статистичної перевірки.

У разі підтвердження стаціонарності похибка повинна бути більш-менш постійною протягом усього музичного твору, це, в свою чергу, є підставою для ідентифікації музичного твору за $D_{ПВ}$, що є МО похибки кластеризації, на основі фрагменту, причому обраному незалежно від проміжку часу.

Отже, в формалізованому вигляді випадкова функція E(t) називається стаціонарною, якщо всі її ймовірнісні

характеристики не залежать від часу t: математичне очікування $m_E(t)$, кореляційна функція $K_E(t,t+r)$ (включає дисперсію $D_E(t)$).

Перевірка стаціонарності процесу здійснювалась для 20 музичних творів. Результати наведені для 4 з них. Проаналізуємо отримані дані з точки зору ймовірної стаціонарності похибки кластеризації E(t). Математичне очікування $m_E(t)$ на фрагментах τ=15с відхиляється від математичного очікування, розрахованого для всього музичного твору $\tilde{M}_E$, $\tilde{m}_{E_{min}} \le \tilde{M}_E \le \tilde{m}_{E_{max}}$, незначно, наприклад: $1 - 0{,}28 \le 0{,}31 \le 0{,}32$; $2 - 0{,}36 \le 0{,}38 \le 0{,}40$; $3 - 0{,}40 \le 0{,}45 \le 0{,}48$; $4 - 0{,}37 \le 0{,}43 \le 0{,}50$. Графіки кореляційної функції (показаної на рисунку 2), отримані для фрагментів з початку, середини і кінця музичного твору (а саме другого, дев'ятого і шістнадцятого фрагментів по 15с), мають подібний характер.

Таким чином, можна говорити про те, що припущення про стаціонарний характер процесу зміни значень похибки кластеризації E(t) підтверджено, відповідно ідентифікацію невідомого музичного твору доцільно здійснювати за його фрагментом.

Слід підкреслити, що особливо важливим є МО похибки M_E , оскільки ця характеристика є визначальною для ідентифікації музичного твору (згідно з формулами (3.1), (3.2)).

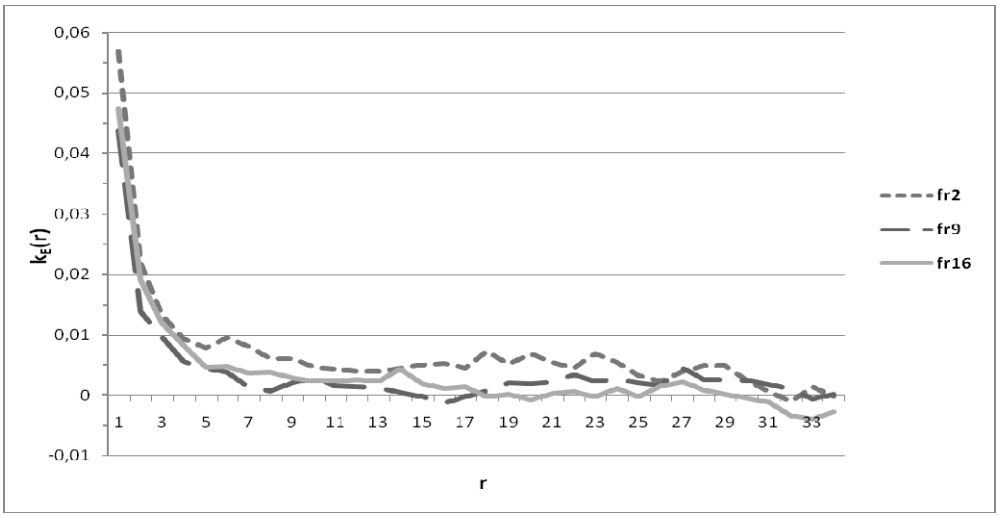

Рисунок 2. Кореляційна функція для другого (fr2), дев'ятого (fr9) і шістнадцятого (fr16) фрагментів по 15с музичного твору

Хоча отримані статистичні дані підтвердили стаціонарність процесу зміни похибки (в результаті кластеризації), значення МО від фрагменту до фрагменту $\tilde{m}_E(t_i), i = \tau, 2\tau, \ldots, T_{song}$ все ж таки дещо коливаються відносно МО в цілому для музичного твору $\tilde{M}_E = D_{ПВ}$ в межах від $\tilde{m}_{E_{min}} = \min(\tilde{m}_E(t_i)), i = \tau, 2\tau, \ldots, T_{song}$ до $\tilde{m}_{E_{max}} = \max(\tilde{m}_E(t_i)), i = \tau, 2\tau, \ldots, T_{song}$:

$\tilde{m}_{E_{min}} \le \tilde{M}_E \le \tilde{m}_{E_{max}}$, або $\tilde{m}_{E_{min}} \le D_{ПВ} \le \tilde{m}_{E_{max}}$.

Виходячи з цього сформулюємо умову, згідно з якою можна ідентифікувати невідомий фрагмент музичного твору, тобто визначити шаблон БД, що є його власним. Фрагмент аудіозапису можна вважати розпізнаним, якщо для відстані між фрагментом та певним шаблоном БД (що відповідно є власним шаблоном фрагмента) виконується система нерівностей:

$$\begin{cases} \tilde{m}_{E_{min}} \le D(\mathbf{X}, \tilde{\mathbf{X}}) \le \tilde{m}_{E_{max}} ; \\ \left| D(\mathbf{X}, \tilde{\mathbf{X}}) - D_{ПВ} \right| \to \min . \end{cases} \qquad (4.1)$$

Відзначимо, що відповідно до формули (4.1), на основі якої має здійснюватись ідентифікація фрагмента музичного твору в БД музики, крім самих шаблонів музичних творів, для кожного музичного твору БД також мають зберігатись значення $\tilde{m}_{E_{min}}, \tilde{m}_{E_{max}}$ та $D_{ПВ}$.

5. Мінімальна і достатня тривалість фрагменту

Крім надійності пошуку, велике значення має швидкість пошуку. Тому важливо, щоб тривалість фрагменту була якомога меншою. Відповідно важливою задачею є визначення мінімальної тривалості фрагменту, що дозволяє ідентифікувати музичний твір за шаблоном. З цією метою було проведено дослідження, аналогічні описаним у розділі 4, на фрагментах тривалістю 1с та 5с.

Нагадуємо, що приведена власна відстань $D_{ПВ}$, на основі якої здійснюється ідентифікація фрагменту музичного твору, є МО похибки кластеризації ($D_{ПВ} = \tilde{M}_E$) відповідно до формул (3.1) та (3.2). На рисунку 3 показано процес зміни в часі значень МО похибки на фрагментах 1с, 5с та 15с. На рисунку 4 показано результати для мінімального $\tilde{m}_{E_{min}}$ та максимального $\tilde{m}_{E_{max}}$ значення МО похибки на фрагментах тривалістю 1, 5, 15с та усереднене значення МО в цілому для музичного твору $\tilde{M}_E$.

Отримані результати свідчать, що для усіх наведених тривалостей τ значення МО від фрагменту до фрагменту

$\tilde{m}_E(t_i), i = \tau, 2\tau, \ldots, T_{song}$ коливаються незначно відносно $\tilde{M}_E$, тобто стаціонарність зберігається.

З рисунку 3 також можна бачити, що зі зменшенням тривалості фрагменту τ коливання $\tilde{m}_E(t_i)$ стають більшими. Зростання коливань $\tilde{m}_E(t_i)$ зі зменшенням τ фрагменту, особливо за умов великої кількості музичних творів в БД, може призвести до виникнення і швидкого зростання кількості випадків неправильного прийняття рішення, коли інший шаблон музичного твору буде прийнятий за власний.

Таким чином, враховуючи отримані результати та зростання рівня коливань для фрагментів тривалістю τ=1с, через недостатню кількість статистичних даних для висновків, було обрано мінімальну тривалість фрагменту – τ=5с для ідентифікації музичного твору за шаблоном.

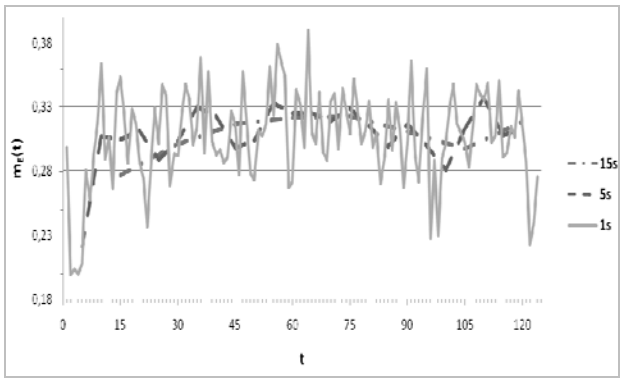

Рисунок 3. Зміна значень МО похибки в часі для 1,5 та 15с

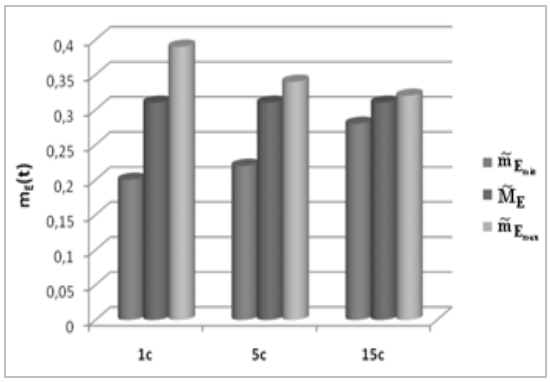

Рисунок 4. Діапазон коливань $\tilde{m}_E(t_i), i = \tau, 2\tau, \ldots, T_{song}$ залежно від τ

6. Перевірка відповідності теоретичних припущень та експериментальних результатів

Для підтвердження теоретичних положень, наведених в попередніх розділах, було проведено експериментальні дослідження на базі 1000 музичних творів. Всі музичні твори мали формат wav (mono), 44,1кГц. В процесі формування БД аудіозаписи шаблонів необхідно було: поділити на фрейми по 20мс з перекриттям 10мс; для кожного фрейму розрахувати вектор параметрів MFCC розмірності 13. Послідовності векторів параметрів MFCC, що описують музичні твори, було кластеризовано, використовуючи вдосконалений метод кластеризації k-середніх, запропонований у одній з попередніх робіт [7]. В результаті чого кожен шаблон БД було представлено 1000 кластерів MFCC. Одночасно в процесі кластеризації для кожного з 1000 музичних творів БД було визначено значення $\tilde{m}_{E_{min}}, \tilde{m}_{E_{max}}$ та $D_{ПВ}$, на основі яких має здійснюватись ідентифікація фрагмента музичного твору в БД музики згідно з формулою (4.1).

З 1000 еталонів БД випадковим чином було обрано фрагменти 100 пісень для їх ідентифікації. Отримані експериментальні результати підтвердили потенційну можливість неправильного прийняття рішення під час ідентифікації фрагменту тривалістю 1с, коли інший шаблон БД приймається за власний.

7. Висновки

В роботі теоретично обґрунтовано можливість ідентифікації музичного твору за його фрагментом на основі приведеної власної відстані (МО похибки кластеризації), значення якої не залежить від кількості фреймів (тривалості запису). Запропоновано аналітичний вираз для визначення власного шаблону музичного твору на основі його фрагменту. Визначено мінімальну тривалість фрагменту (5с), що дозволяє зменшити складність обчислень в десятки разів в процесі автоматичної ідентифікації музичного твору. Проведені експериментальні дослідження підтвердили справедливість наведених теоретичних положень.

8. Список літератури

[1] Ganchev T. Comparative evaluation of various mfcc implementations on the speaker verification task / T. Ganchev, N. Fakotakis, and G. Kokkinakis // Proceedings of 9th International Conference on Speech and Computer, SPECOM'05. – 2005. – pp. 191–194.

[2] Logan B. A music similarity function based on signal analysis / B. Logan and A. Salomon // Proc. IEEE Int. Conf. Multimedia Expo. – 2001. – pp. 745–748.

[3] Tzanetakis G. Musical genre classification of audio signals / G. Tzanetakis and P. Cook // IEEE Trans. Speech Audio Process. – No. 5. – V. 10. – 2002. – pp. 293–301.

[4] Senin P. Dynamic time warping algorithm review / P. Senin – Honolulu, USA. – 2008.

[5] Gersho A. Vector Quantization and Signal Compression. / A. Gersho, R. M. Gray. – Boston: Kluwer Academic. – 1992. – 760 p.

[6] Jain A. K. Algorithms for Clustering Data / A. K. Jain, R. C. Dubes. – Englewood Cliffs, N.J.: Prentice Hall. – 1988. – 334 p.

[7] Ткаченко О.М. Метод кластеризації на основі послідовного запуску k-середніх з удосконаленим вибором кандидата на нову позицію вставки / О. М. Ткаченко, О. Ф. Грійо Тукало, О. В. Дзісь, С. М. Лаховець // Електронний журнал «Наукові праці ВНТУ». – №2. – В.2. – Вінниця: ВНТУ. – 2012.