

УДК 621.39

О. М. Ткаченко, канд. тех. наук, доц.; І. Р. Арсенюк, канд. тех. наук, доц.;
С. В. Хрушак, канд. тех. наук

МЕТОД ШВИДКОГО ПОШУКУ ФРАГМЕНТА АУДІОЗАПИСУ

Враховуючи великі обсяги аудіоінформації, що зберігається у корпусі музичних творів, велике значення має швидкість та надійність їх пошуку. В статті пропонується метод швидкого пошуку фрагмента аудіозапису з удосконаленою оцінкою ступеня близькості між невідомим аудіофрагментом і шаблонами, який дозволяє підвищити достовірність прийняття рішення під час пошуку.

Для компактного опису параметрів сигналу обрано мел-частотні кепстральні коефіцієнти, на основі яких формується корпус параметрів музичних творів як набір центроїдів, отриманих у результаті кластеризації. Введено поняття власної приведенної відстані як оцінки ступеня близькості невідомого фрагмента музичного твору та попередньо створених шаблонів аудіозаписів. Обґрунтовано застосування kd-дерев для прискорення пошуку невідомого фрагменту в корпусі аудіозаписів та наведено основні етапи пошуку. Розглянуто різні варіанти обчислення міри близькості невідомого аудіофрагмента з аудіозаписами у корпусі, а саме: оцінювання міри близькості за приведеною відстанню, оцінювання міри близькості за кількістю попадань у список k найближчих центроїдів, оцінювання міри близькості за зваженою кількістю попадань у список k найближчих центроїдів. Показано, що виконання не точного, а наближеного пошуку векторів на основі kd-дерева дозволяє досягти значної економії часу, проте призводить до зниження достовірності результатів пошуку. Тому для зменшення складності обчислень при збереженні достовірності результатів запропоновано для великих архівів аудіозаписів виконувати комбінований пошук, що поєднує швидкий "неточний" пошук із застосуванням kd-дерева кількох найближчих аудіозаписів корпусу для заданого аудіофрагменту на першому етапі пошуку, серед яких на другому етапі за допомогою повного перебору визначається один найближчий. Запропонований метод дозволив підвищити повноту та релевантність результатів пошуку.

Ключові слова: аудіозапис, корпус музичних творів, міра близькості, обчислення відстані, кластеризація, kd-дерева.

Вступ

Зростання обсягів мультимедійної інформації, зокрема аудіоінформації, яка передається та обробляється у комп'ютерних системах, зумовила необхідність автоматизації процесів аналізу та пошуку даних. Тому в сучасних системах обробки аудіоінформації виникає необхідність автоматичного пошуку музичних творів (МТ) на основі аудіоконтенту в базах даних. Враховуючи великі обсяги аудіоінформації, що зберігається у корпусі МТ, велике значення має швидкість та надійність пошуку. Для підвищення надійності результатів пошуку в статті пропонується метод швидкого пошуку фрагмента аудіозапису в корпусі МТ з удосконаленою оцінкою ступеня близькості між невідомим аудіофрагментом і шаблонами, що дозволяє підвищити достовірність прийняття рішення під час пошуку за рахунок збільшення ступеня розбіжності власного МТ з іншими аудіозаписами.

Мета та задачі статті

Метою цієї статті є підвищення достовірності прийняття рішення у процесі пошуку шаблону власного МТ (далі у статті – власного шаблону) для невідомого аудіофрагменту завдяки застосуванню удосконаленої оцінки ступеня близькості між фрагментом та шаблонами, що дозволяє істотно збільшити рівень розбіжності власного шаблону з іншими.

Для підвищення достовірності прийняття рішення під час пошуку власного шаблону для невідомого фрагмента необхідно вирішити такі задачі:

1. Вибрати параметри, які б однозначно і компактно його описували.

2. Запропонувати оцінку ступеня близькості невідомого фрагмента МТ та попередньо створених шаблонів аудіозаписів, що забезпечать високий рівень розбіжності власного шаблону та інших.

3. Провести експериментальні дослідження.

Результатом пошуку буде шаблон МТ у корпусі з мінімальною відмінністю з вхідним аудіозаписом. Далі у роботі вважається, що аудіозапис, який треба визначити, точно збігається з одним із шаблонів корпусу.

Вибір математичної моделі аудіосигналу

Схема визначення відповідності МТ на основі контенту базується на використанні аудіофайлу для побудови моделі аудіосигналу. Порівнювати безпосередньо звукові сигнали у часовій області неефективно, тому аудіозапис розбивають на фрейми (фрагменти по 10 – 30 мс), протягом яких характеристики сигналу залишаються відносно стійкими. Для кожного фрейму виконується спектральний аналіз, на основі якого обчислюється значення вектора параметрів (параметризація). Серед різних параметрів обрано мел-частотні кепстральні коефіцієнти (MFCC – Mel Frequency Cepstral Coefficients), які вперше було запропоновано використовувати у системах розпізнавання мови і диктора [1], надалі вони отримали широке застосування у процесі інформаційного пошуку музики (Music information retrieval, MIR) [2]. У результаті параметризації отримуємо опис МТ як файл, що містить вектори MFCC.

Отже, якщо частота дискретизації звукового файлу – 44,1кГц, довжина фрейму – 20 мс з перекриттям 0,5 кадру, то середньостатистичний МТ тривалістю 3 хв. характеризують приблизно 8 млн. відліків, або, з урахуванням перекриття кадрів, приблизно 18 тис. кадрів, кожен з яких описується вектором параметрів MFCC розмірності 13 (рис. 1).

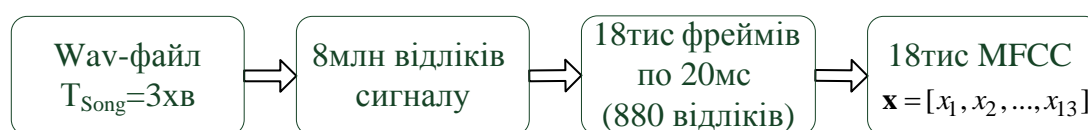


Рис. 1. Опис wav-файлу (3 хв.)

Як бачимо, параметризація дозволяє зменшити кількість інформації, необхідної для опису МТ, у десятки разів:

$$\frac{8\,000\,000\text{ відліків}}{18\,000\text{ MFCC} \cdot 13} = \frac{8\,000\,000}{234\,000} \approx 34\text{ рази}$$

Таким чином, коефіцієнти MFCC є компактним поданням спектральної огинаючої, що при розпізнаванні МТ дозволяє успішно замінити мільйони відліків аудіофайлу.

Порівняння невідомого МТ з шаблонами за приведеною відстанню

Після вибору параметрів для опису МТ необхідно перейти до порівняння невідомого МТ з аудіозаписами корпусу та визначення шаблону, відмінність з яким буде мінімальною. Щоб ідентифікувати невідомий аудіозапис, необхідно мати критерій порівняння. Як правило, таким критерієм є відстань. Порівняння за вибраними параметрами має забезпечити високий рівень розбіжності власного шаблону та шаблонів інших творів. Тобто у результаті порівняння невідомого МТ із власним шаблоном похибка (по відстані) між ними має бути мінімальною, і якнайбільшою для шаблонів інших творів:

$$\begin{cases} D(X, \tilde{X}) \rightarrow \min \\ D(X, \tilde{Y}) \rightarrow \max \end{cases} \Rightarrow D(X, \tilde{X}) \ll D(X, \tilde{Y}), \quad (1)$$

де \tilde{X} – множина векторів параметрів власного шаблону; \tilde{Y} – множина векторів параметрів шаблону іншого МТ.

Загальну схему пошуку МТ наведено на рис. 2.

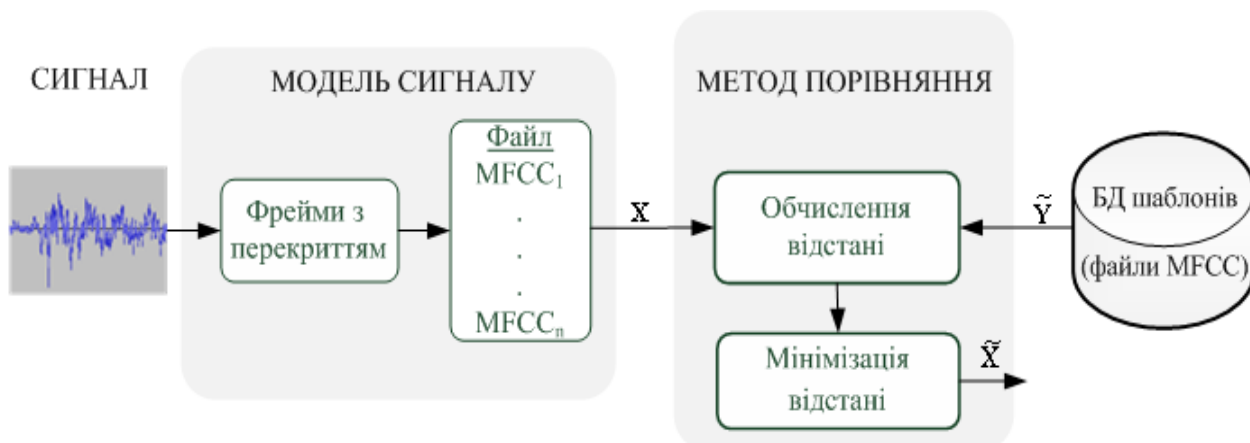


Рис. 2. Загальна схема пошуку МТ

Найпростішим та очевидним підходом для визначення близькості між наборами параметрів MFCC невідомого МТ та аудіозаписів у корпусі є порівняння MFCC на основі найбільш поширеної Евклідової метрики, точніше квадрата Евклідової відстані. Формула незваженої Евклідової відстані між вектором параметрів $x = (x_1, x_2, \dots, x_d)$ МТ, який треба ідентифікувати, і вектором шаблону $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_d)$:

$$D_{Eu}^2(x, \tilde{y}) = \sum_{i=1}^d (x_i - \tilde{y}_i)^2. \quad (2)$$

Відповідно відстань (похибку) між файлами параметрів невідомого твору та шаблону можна знайти як суму відстаней:

$$D(X, \tilde{Y}) = \sum D_{Eu}^2(x_j, \tilde{y}_j). \quad (3)$$

В ідеальному випадку при такому підході відстань між файлами параметрів MFCC одного і того ж МТ дорівнюватиме нулю, для різних – відмінною від нуля:

$$\begin{cases} D(X, \tilde{X}) = 0, \\ D(X, \tilde{Y}) > 0. \end{cases} \quad (4)$$

Однак слід зауважити, що навіть для одного й того ж МТ аудіозаписи можуть відрізнятися, наприклад, на початку запису може бути тиша, мелодія іншого МТ тощо; записи можуть мати різний темп, тривалість. Це означає, що при зміщенні кадрів у часі відстань до власного шаблону $D(X, \tilde{X}) \neq 0$, тобто умова розрізнення власного МТ (4) не виконується, отже, безпосереднє порівняння файлів параметрів за Евклідовою відстанню не підходить для пошуку аудіофрагменту в корпусі аудіозаписів.

Очевидно, що у більшості випадків МТ характеризується певною періодичністю, що полягає у наявності ідентичних або дуже схожих за текстом та характером мелодії фрагментів. Відповідно можна говорити про надмірність даних, якими описується аудіозапис, та скоротити кількість параметрів для його опису. Враховуючи це, доцільно застосувати методи кластерного аналізу для формування корпусу аудіозаписів. Використання кластеризації для формування еталонів корпусу дозволить зменшити обсяги пам'яті, необхідні для їх зберігання.

Основні етапи порівняння файлів параметрів невідомого МТ із певним шаблоном у корпусі:

1. Пошук мінімальної Евклідової відстані D_{min}^2 між поточним вектором параметрів $\mathbf{x} = (x_1, x_2, \dots, x_d)$ із множини параметрів $\mathbf{X} = \{\mathbf{x}\}, |\mathbf{X}| = n$ МТ, який слід ідентифікувати, та множиною векторів-центроїдів, $\tilde{\mathbf{Y}} = \{\tilde{\mathbf{y}}\}, |\tilde{\mathbf{Y}}| = m$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_d)$ шаблону:

$$D_{min}^2 = \min_j (D_{Eu}^2(\mathbf{x}, \tilde{\mathbf{y}}_j)) = \min_j \left(\sum_{i=1}^d (x_i - \tilde{y}_{ji})^2 \right), j = \overline{1, m}. \quad (5)$$

2. Обчислення оцінки відстані в цілому до шаблону у вигляді суми квадратів мінімальних відстаней D_{min}^2 :

$$D(\mathbf{X}, \tilde{\mathbf{Y}}) = \sum_{l=1}^n D_{min}^2 = \sum_{l=1}^n \min_j \left(\sum_{i=1}^d (x_i - \tilde{y}_{ji})^2 \right), j = \overline{1, m}. \quad (6)$$

Разом з тим, оскільки МТ мають різну тривалість, кожен твір характеризується власною кількістю фреймів, представлених векторами MFCC. Під час кластеризації обчислюється однакова кількість центроїдів усіх шаблонів. Це призводить до того, що різні твори знаходяться у нерівних умовах, тобто для творів, тривалість яких більша, початкова похибка (між файлами того ж твору до і після кластеризації) теж буде більшою, оскільки у цьому випадку на кожен кластер припадатиме більше параметрів векторів. Уникнути цього можна, якщо поділити відстань до шаблону, визначену у формулі (6), на кількість фреймів МТ. Назвемо цю величину приведеною відстанню (ПВ) – $D_{ПВ}$ аудіозапису:

$$D_{ПВ} = \frac{D(\mathbf{X}, \tilde{\mathbf{Y}})}{n} = \frac{\sum_{l=1}^n D_{min}^2}{n} = \frac{\sum_{l=1}^n \min_j \left(\sum_{i=1}^d (x_i - \tilde{y}_{ji})^2 \right)}{n}, j = \overline{1, m}. \quad (7)$$

Як бачимо з формули (7) характеристика $D_{ПВ}$ не залежить від кількості кадрів (тривалість запису). Отже, її можна використовувати як критерій прийняття рішення як для запису в цілому, так і для окремого аудіофрагменту. Обґрунтування цього наведено у [3], там же визначено тривалість фрагмента МП у 5 с (500 кадрів, або 500 векторів параметрів MFCC), достатня для пошуку аудіозапису.

Створення корпусу шаблонів МТ на основі кластеризації

Методи кластеризації широко використовують у системах розпізнавання аудіосигналу. Завдання кластеризації можна сформулювати так: заданий набір з n векторів розмірності d необхідно розбити на підмножини відповідно до критерію мінімізації спотворення $e_i^2 \rightarrow \min$. Існують різні шляхи оцінювання спотворення, але у більшості прикладних реалізацій використовують суму середньоквадратичних Евклідових відстаней між центром кластера (центроїдом) \mathbf{c}_i і векторами параметрів, які йому належать $\mathbf{X}_i = \{\mathbf{x}\}, \mathbf{X}_i \subset \mathbf{X}$ [4], тобто:

$$e_i^2 = \left\{ \mathbf{c}_i : \sum_{j=1}^{N_i} D_{Eu}^2(\mathbf{x}_j, \mathbf{c}_i) / \mathbf{x} \in \mathbf{X}_i \leq \sum_{j=1}^{N_i} D_{Eu}^2(\mathbf{x}_j, \mathbf{c}) / \mathbf{x} \in \mathbf{X}_i \right\};$$

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

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

Таким чином, шаблони МТ можна описати центроїдами кластерів параметрів MFCC. Оскільки між кожним центроїдом \mathbf{c}_i і векторами, що йому належать, є похибка e_i^2 , то відстань між файлами параметрів, що описують той самий МТ до кластеризації і після (навіть якщо аудіозаписи були ідентичними), буде позитивним і рівним величині сумарної похибки кластеризації E^2 .

Отже, сформований корпус шаблонів МТ для ідентифікації невідомого фрагмента аудіозапису фактично є файлом, що містить центроїди MFCC всіх шаблонів МТ.

Застосування kd-дерев для пошуку фрагмента МТ у корпусі шаблонів

Найпростіше рішення загальної задачі пошуку найближчого сусіда, а у нашому випадку – найближчого центроїду в корпусі аудіозаписів – це обчислення відстані між вектором невідомого аудіофрагменту, який потрібно ідентифікувати, та усіма векторами-центроїдами шаблонів. Час виконання повного пошуку пропорційно $O(dN)$, де N – кількість центроїдів у корпусі, d – розмірність векторів. Однак застосувати такий підхід для пошуку у великих базах даних практично неможливо. Тому для зменшення складності обчислень під час пошуку фрагмента аудіозапису у сформованому корпусі був застосований простий спосіб розбиття простору – k -вимірне дерево (kd-дерево). Центроїди аудіозаписів у корпусі були упорядковані з урахуванням kd-дерева [5]. Упорядкування виконується дуже швидко, оскільки розбиття простору відбувається лише уздовж осей даних, відповідно при цьому не потрібно обчислювати d -вимірні відстані.

Щоб забезпечити отримання найближчого вектора, пошук, крім прямої фази (спуску по дереву), повинен мати і зворотну (пошук елементів дерева у заданому діапазоні). Під час прямого пошуку фіксуються усі відстані до вузлів \mathbf{d}_i . Пряма фаза завершується обчисленням відстані $D_{min} = D_k$ до відповідної термінальної вершини, в якій лежить невідомий вектор, а також потенційно найближчий елемент дерева (центроїд шаблону в корпусі), що задає зворотній фазі радіус пошуку. Під час зворотної фази пошуку обчислюються відстані \mathbf{d}_i лише до тих вузлів дерева, які можуть забезпечити виконання $d_i < D_{min}$. Якщо для відповідного листа виконується умова $D_k < D_{min}$, радіус пошуку корегують $D_{min} = D_k$ [6].

Розглянута процедура пошуку гарантує знаходження найближчого вектора за формулою (1), проте потребує більшої кількості вимірювань відстані, ніж $\log_2 n$.

Зазначимо, що пошук на основі kd-дерева є дуже ефективним лише для невеликих розмірностей ($d < 20$). Для $d > 20$ середній час пошуку найближчого вектора у корпусі, впорядкованого на основі kd-дерева, у середньому зростає пропорційно $O(h) = O(d \cdot \log(N+1))$, де h – висота дерева [6]. Для більших розмірностей обчислювальна складність може досягати $O(d \cdot N)$.

Отже, послідовність кроків під час виконання пошуку відповідного для аудіофрагмента аудіозапису в корпусі на основі kd-дерева така.

1. Виконується швидкий пошук в упорядкованій на основі kd-дерева базі параметрів корпусу аудіозаписів, у процесі якого за Евклідовою метрикою, згідно з формулою (2), відбирається множина \mathbf{C} векторів-центроїдів аудіозаписів (кандидатів на найближчий вектор), упорядкованих за зростанням відстані:

$$\mathbf{C} \subset \mathbf{T}, \mathbf{C} = \{\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \dots, \tilde{\mathbf{Y}}_k\}, |\tilde{\mathbf{Y}}| = m, |\mathbf{C}| = k, |\mathbf{T}| = N = m \cdot Q, k \leq N,$$

де k – кількість векторів аудіозаписів у списку найближчих до фрейму невідомого аудіофрагмента; m – кількість центроїдів (векторів) аудіозапису; Q – загальна кількість аудіозаписів у корпусі; N – загальна кількість центроїдів-векторів корпусу аудіозаписів.

Для кожного з k найближчих центроїдів множини C зберігаються такі дані: відстань до фрейму невідомого аудіофрагмента $D_p^2(\mathbf{x}_l, \tilde{\mathbf{y}}_{qj})$ та індекс відповідного аудіозапису i_{pq} у корпусі:

$$\mathbf{I} = \{i_{pq}\}, \mathbf{D} = \{D_p^2(\mathbf{x}_l, \tilde{\mathbf{y}}_{qj})\},$$

$$D_p^2(\mathbf{x}_l, \tilde{\mathbf{y}}_{qj}) = \sum_{i=1}^d (x_i - \tilde{y}_i)^2, p = \overline{1, k}, l = \overline{1, n}, j = \overline{1, m}, q = \overline{1, Q}.$$

2. Визначення для аудіофрагмента власного аудіозапису в корпусі. Відстань до кожного з аудіозаписів у корпусі визначається згідно з правилом:

$$Distance_q = \begin{cases} \min(D_p^2), \forall q \in \mathbf{I}; \\ D_p^2, p = k, \forall q \notin \mathbf{I}. \end{cases} \quad (8)$$

У процесі виконання пунктів 1 і 2 обчислюється сума відстаней по усіх n фреймах невідомого аудіофрагмента для кожного аудіозапису $\tilde{\mathbf{Y}}_q$ у корпусі

$$D(\mathbf{X}, \tilde{\mathbf{Y}}_q) = \sum_{l=1}^n Distance_{q_l}. \quad (9)$$

На останньому кроці знаходиться мінімум приведеної відстані, визначеної у формулі (7):

$$D(\mathbf{X}, \tilde{\mathbf{X}}) = \min_q D_{PBq} = \min_q \frac{D(\mathbf{X}, \tilde{\mathbf{Y}}_q)}{n}. \quad (10)$$

Відповідно власний аудіозапис $\tilde{\mathbf{X}}$ у корпусі як найближчий до невідомого аудіофрагмента вважається визначеним.

Наведена процедура пошуку забезпечує отримання не одного найближчого вектора, а деякої множини векторів, упорядкованих за зростанням відстані згідно з формулою (7). Це можливо за рахунок того, що дані про відстань до пройдених термінальних вершин зберігаються, і одночасно здійснюється їх упорядкування за зростанням відстані до невідомого вектора. Завдяки цьому додаткове знаходження кількох найближчих векторів не потребує багато часу.

Метод швидкого пошуку фрагмента аудіозапису з різними варіантами оцінки міри близькості

З метою підвищення надійності результатів пошуку за рахунок збільшення міри розрізнення між власним аудіозаписом та іншими запропоновано можливі варіанти обчислення міри близькості невідомого аудіофрагмента з аудіозаписами у корпусі, тобто варіанти виконання другого кроку пошуку на основі kd-дерева.

Оцінка міри близькості за приведеною відстанню

1. Для поточного фрейму \mathbf{x}_l невідомого аудіофрагмента відстань до кожного аудіозапису $\tilde{\mathbf{y}}_q$ у корпусі визначається згідно з правилом, вираженим формулою (8).

Тобто, на цьому фреймі відстанню до аудіозапису буде вважатися найменша відстань серед його центроїдів, що потрапили у список k найближчих. Відстань для аудіозаписів, центроїди яких не потрапили у список k найближчих, вважається такою, що дорівнює відстані до останнього k -го центроїда зі списку найближчих. Останнє фактично означає, що

ми штучно зменшуємо відстань до тих аудіозаписів, що не потрапили у список k найближчих, і, відповідно, зменшуємо рівень розрізнення власного аудіозапису та решти аудіозаписів у корпусі.

2. Для кожного аудіозапису \tilde{Y}_q у корпусі обчислюється сума відстаней по усіх n фреймах невідомого аудіофрагмента згідно з (9).

3. Власний аудіозапис у корпусі визначається з мінімуму приведеної відстані згідно з формулою (10).

З метою збільшення різниці відстаней між власним аудіозаписом та рештою аудіозаписів корпусу нижче запропоновано інші міри оцінювання близькості аудіозаписів.

Оцінка міри близькості за кількістю попадань у список k найближчих центроїдів (Hits)

1. Для кожного фрейму x_l невідомого аудіофрагмента фіксується кількість попадань центроїдів q -го аудіозапису в список k найближчих:

$$\mathbf{Hits}_q = \sum_{p=1}^k h_p, \text{ де } h_p = \begin{cases} 1, \forall q \in \mathbf{I}; \\ 0, \forall q \notin \mathbf{I}. \end{cases} \quad (11)$$

2. Для кожного аудіозапису \tilde{Y}_q знаходиться сумарне значення кількості попадань по усіх n фреймах невідомого аудіофрагмента.

3. Власним аудіозаписом вважається той, кількість попадань по усіх n фреймах для якого є найбільшою:

$$H(X, \tilde{X}) = \max_q \sum_{l=1}^n \mathbf{Hits}_q. \quad (12)$$

При такому підході різниця відстаней між власним аудіозаписом та іншими зростає, також зменшується складність обчислень.

Оцінка міри близькості за зваженою кількістю попадань у список k найближчих центроїдів (Weighted Hits)

Цей підхід до оцінювання міри близькості між невідомим аудіофрагментом та базою параметрів аудіозаписів корпусу є комбінацією двох попередніх, тобто враховуються відстані D_p^2 , кількість попадань h_p центроїдів аудіозапису в список k найближчих, а також позиція аудіозапису в списку найближчих, що відображено у формулі (13).

1. Для кожного фрейму x_l невідомого аудіофрагмента обчислюється величина, яку було названо зваженою кількістю попадань центроїдів q -го аудіозапису в список k найближчих, за формулою:

$$\mathbf{Weighted Hits}_q = \sum_{p=1}^k (D_k^2 - D_p^2) \cdot h_p. \quad (13)$$

2. Аналогічним чином для кожного аудіозапису \tilde{Y}_q знаходиться сумарне значення зваженої кількості попадань по усіх n фреймах невідомого аудіофрагмента.

3. Власним аудіозаписом вважається той, значення зваженої кількості попадань по усіх n фреймах для якого є найбільшим, причому поділене на загальну кількість фреймів:

$$\mathbf{WH}(X, \tilde{X}) = \max_q \sum_{l=1}^n \frac{\mathbf{Weighted Hits}_q}{n}. \quad (14)$$

Цей удосконалений підхід до оцінювання міри близькості між невідомим

аудіофрагментом та аудіозаписами корпусу має врівноважити переваги і недоліки попередніх двох.

Отже, запропоновано метод пошуку власного аудіозапису з удосконаленою оцінкою міри близькості. Наведемо послідовність дій для його виконання.

Для кожного фрейму (вектора) x_i невідомого аудіофрагмента виконується нижчезказане:

1. Швидкий пошук k найближчих сусідів у кластеризованому корпусі аудіозаписів, упорядкованому на основі kd-дерева.
2. Визначення для невідомого аудіофрагмента його власного аудіозапису в корпусі за зваженою кількістю потраплянь у список k найближчих центроїдів.

Експериментальні результати

Для проведення експериментальних досліджень було сформовано корпус із 1000 аудіозаписів МТ. Усі МТ мали формат wav (mono) із частотою дискретизації 44,1кГц. Попередньо з аудіофайлів було видалено «тишу» на початку та у кінці записів. У процесі формування корпусу МТ аудіозаписи шаблонів було розбито на фрейми по 20 мс із перекриттям 10 мс і для кожного кадру обчислено вектор параметрів MFCC розмірності 13. Послідовності векторів параметрів MFCC, що описують МТ, було кластеризовано, використовуючи удосконалений метод кластеризації [7]. Після кластеризації кожен шаблон у корпусі було представлено 1000 центроїдами MFCC (загалом близько 20 векторів на кластер). Фрагменти тривалістю 5 с для ідентифікації вибиралися випадково з еталонів сформованого корпусу МП. Застосування пошуку по kd-дереву та оцінки міри близькості до шаблонів порівняно з повним пошуком призводить до деякого зменшення ступеня розрізнення власного аудіозапису та інших шаблонів корпусу (наведено на рис. 3), проте різниця залишається достатньою для прийняття правильного рішення.

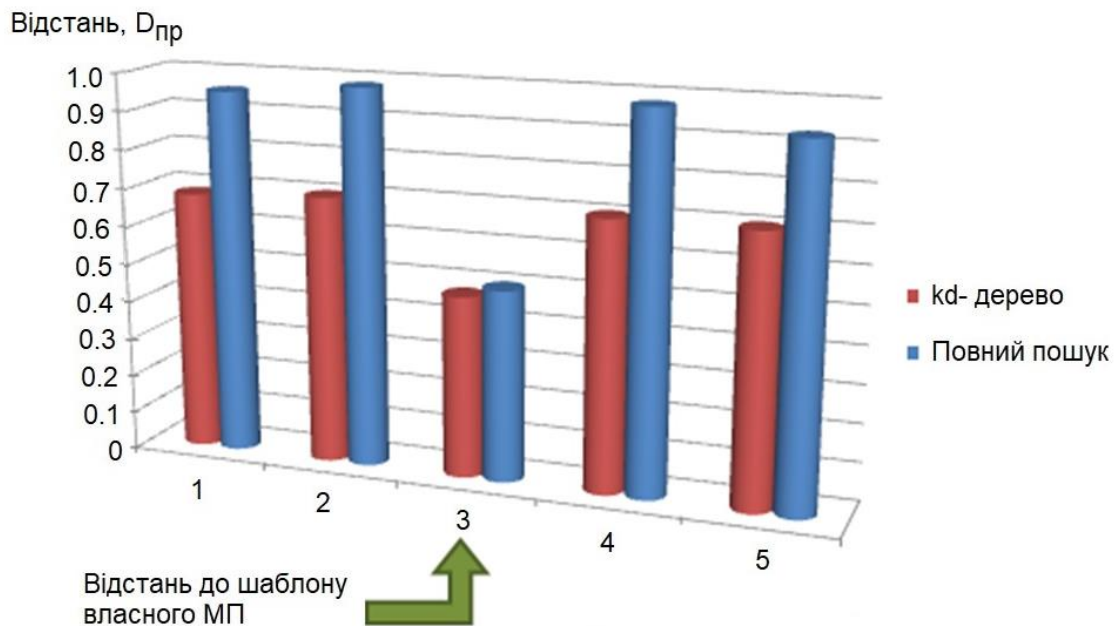


Рис. 3. Порівняння результатів повного пошуку та пошуку на основі kd-дерева

На рис. 4 наведено результати порівняння невідомого фрагмента із шаблонами (власний шаблон за №3) для оцінки ступеня їх близькості, та зведену гістограму, на якій представлено всі варіанти оцінювання ступеня близькості шаблонів та невідомого фрагмента у відносних одиницях задля досягнення єдиного масштабу.

Як бачимо з наведених результатів, оцінювання ступеня близькості за зваженою кількістю

влучань шаблону до списку k найближчих (Weighted Hits) забезпечує високий рівень розрізнення власного та чужих шаблонів.

Також на рівень розрізнення власного шаблону та інших впливає кількість центроїдів, якими представлено аудіозаписи у корпусі, тривалість фрагмента для пошуку та кількість найближчих центроїдів, які отримуються в результаті пошуку по kd -дереву. Останній варіант є найбільш прийнятним для підвищення рівня розрізнення, тому що не призводить до значного зростання часу пошуку. Розглядалися результати для 20 та 50 найближчих кандидатів. У таблиці 1 для п'яти фрагментів представлено результати, отримані під час вибору власного аудіозапису в корпусі за оцінками ступеня близькості невідомого фрагмента шаблонів. Для кожної з оцінок ступеня близькості наведено значення, отримані для власного шаблону (виділені кольором), та найближчого шаблону серед інших.

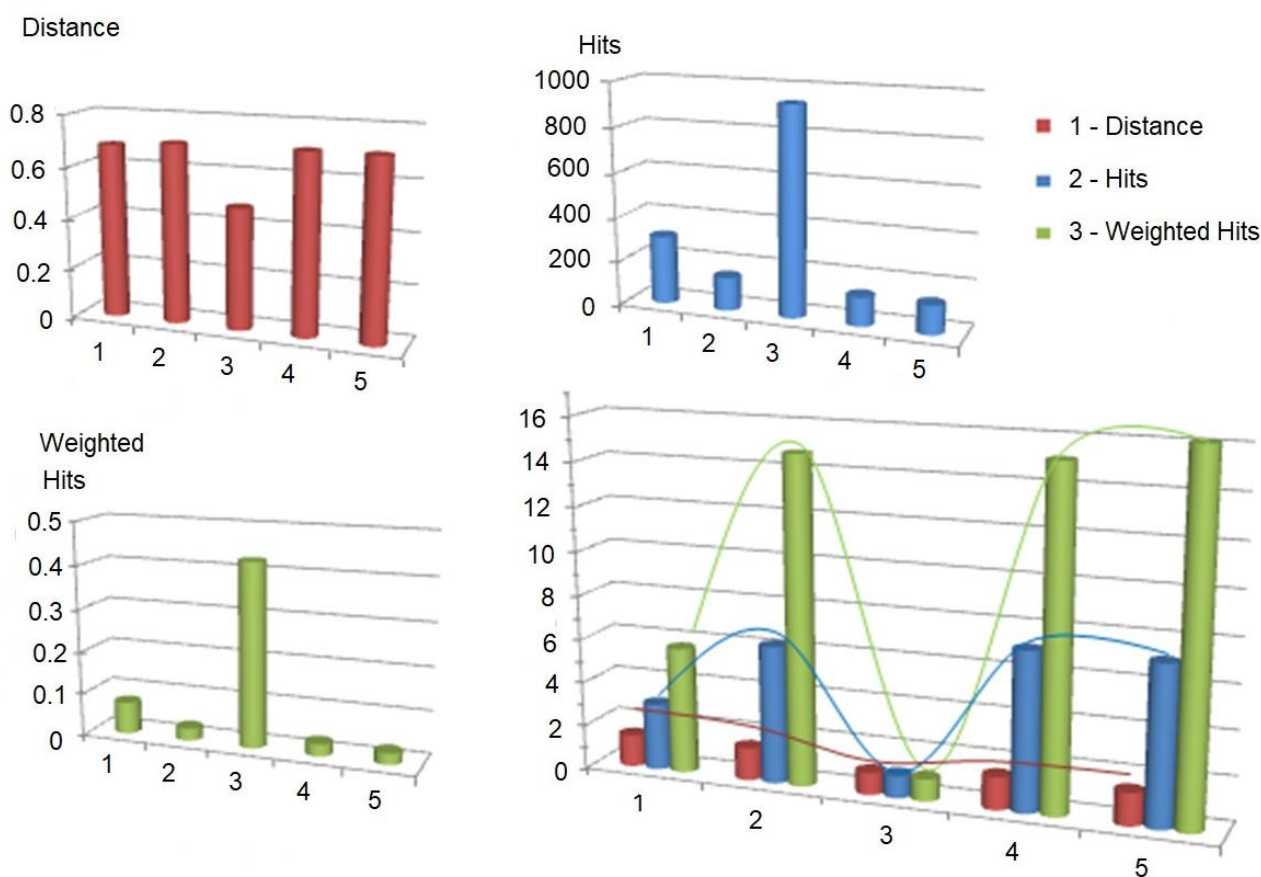


Рис. 4. Порівняння оцінок ступеню близькості невідомого фрагмента з шаблонами

Таблиця 1

Результати запропонованих оцінок ступеню близькості для власного шаблону та найближчого серед інших для $k = 20$ та $k = 50$

k	20					50				
№ фрагм.	1	2	3	4	5	1	2	3	4	5
D_{PB}	0,412	0,454	0,304	0,364	0,454	0,425	0,490	0,324	0,386	0,475
	0,753	0,564	0,434	0,458	0,615	0,847	0,639	0,498	0,521	0,683
H	821	275	391	335	580	1238	412	535	479	931
	160	57	65	74	140	390	134	125	177	307
WH	0,567	0,152	0,174	0,118	0,260	0,900	0,251	0,261	0,191	0,426
	0,075	0,017	0,026	0,016	0,027	0,167	0,048	0,056	0,041	0,074

Отримані дані показують, що зі збільшенням кількості найближчих сусідів до 50 різниця в оцінках зростає, що збільшує ступінь розрізнення між власним шаблоном та іншими.

Оцінка відстані до власного аудіозапису, отримана за формулою (14), при $k = 20$ у середньому у 8 разів перевищує аналогічну оцінку відстані до інших шаблонів. Таким чином, удосконалена оцінка близькості значно підвищує достовірність прийняття рішення у процесі пошуку власного шаблону МП і забезпечує можливість для подальшого розширення корпусу МП.

Насамкінець також зазначимо, що для пошуку в реальних архівах аудіозаписів, що містять сотні тисяч – мільйони аудіозаписів, критичною вимогою є висока швидкість пошуку. Виконання не точного, а наближеного пошуку векторів (коли знайдений вектор не обов'язково найближчий, але досить близький) на основі kd-дерева дозволяє досягти значної економії часу, проте призводить до зниження достовірності результатів пошуку. Тому для зменшення складності обчислень при збереженні достовірності результатів доцільно для великих архівів аудіозаписів виконувати комбінований пошук, що поєднує швидкий "неточний" пошук із застосуванням kd-дерева кількох найближчих аудіозаписів корпусу для заданого аудіофрагменту на першому етапі пошуку, серед яких на другому етапі за допомогою повного перебору визначається один найближчий.

Отримав подальшого розвитку метод пошуку на основі kd-дерева з удосконаленою оцінкою міри близькості за зваженою кількістю попадань у список найближчих центроїдів. Запропонований метод дозволив підвищити повноту та релевантність результатів пошуку, а саме: для аудіофрагменту тривалістю 5 с повнота пошуку становить 99,8 %, релевантність – 97,4 %, для аудіофрагменту 1 с – 99,6 % та 94,3 % відповідно, що перевищує аналогічні показники аналогічних методів [8].

Висновки

У статті запропоновано метод швидкого пошуку аудіофрагментів тривалістю 5 с на основі kd-дерева з удосконаленою оцінкою ступеня близькості між невідомим фрагментом та аудіозаписами у корпусі, що дозволяє підвищити достовірність результатів пошуку. Запропоновано аналітичні співвідношення для оцінки ступеня близькості невідомого фрагмента з шаблонами, що базуються на обчисленні відстані від фрагмента до заданої кількості найближчих центроїдів та кількості попадань до списку найближчих сусідів. За рахунок застосування удосконаленої оцінки ступеня близькості (за зваженою кількістю влучень до списку k найближчих центроїдів) між невідомим аудіофрагментом та шаблонами суттєво збільшено рівень розрізнення власного аудіозапису в корпусі. Експериментальні результати показують, що оцінка відстані до власного аудіозапису при $k = 20$ у середньому у 8 разів перевищує аналогічну оцінку відстані до інших аудіозаписів корпусу.

СПИСОК ЛІТЕРАТУРИ

1. Ganchev T. Comparative evaluation of various mfcc implementations on the speaker verification task / T. Ganchev, N. Fakotakis, G. Kokkinakis // In Proceedings of 9th International Conference on Speech and Computer, SPECOM'05. – 2005. – P. 191 – 194.
2. Li T. Visual Classification of Music Style Transfer Based on PSO-BP Rating Prediction Model / T. Li // Complexity. – 2021. – Vol. 2021. – P. 1 – 9. – URL: <https://doi.org/10.1155/2021/9959082>.
3. Ткаченко О. М. Підхід до оцінювання тривалості фрагмента для пошуку музичного твору за заданим шаблоном / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 1. – С. 31 – 40.
4. Introduction to K-D Trees [Electronic resource] / H. Hristov // Baeldung. – Access mode : <https://www.baeldung.com/cs/k-d-trees> (date of access: 10.02.2024).
5. Moore A. W. Efficient memory based learning for robot control / A. W. Moore. – PhD thesis Nr. – University of Cambridge, Computer Laboratory UCAM-CL-TR-209. – 1990. – 248 p.

6. Jo J. A progressive k-d tree for approximate k-nearest neighbors / J. Jo, J. Seo, J. Fekete // 2017 IEEE Workshop on Data Systems for Interactive Analysis (DSIA), Phoenix, AZ, USA. – 2017. – P. 1 – 5. DOI: 10.1109/DSIA.2017.8339084.

7. Метод кластеризації на основі послідовного запуску k-середніх з обчисленням відстаней до активних центроїдів / О. М. Ткаченко, О. Ф. Грійо Тукало, Н. О. Біліченко [та ін.] // Реєстрація, зберігання і обробка даних. – 2012. – № 1, Т. 14, В. 1. – С. 25 – 34.

8. Ткаченко О. М. Метод підвищення швидкості пошуку фрагменту аудіозапису із застосуванням kd-дерев / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 3. – С. 57 – 66.

Стаття надійшла до редакції 25.02.2024.

Стаття пройшла рецензування 14.03.2024.

Ткаченко Олександр Миколайович – канд. тех. наук, доцент кафедри програмного забезпечення.

Арсенюк Ігор Ростиславович – канд. тех. наук, доцент кафедри комп'ютерних наук, e-mail: igrosars@gmail.com.

Вінницький національний технічний університет.

Хрущак Сергій Вікторович – канд. тех. наук, старший викладач кафедри комп'ютерних наук.

Вінницький національний аграрний університет.