

А.Н. Ткаченко, О.Ф. Грийо Тукало

Метод быстрого поиска фрагмента аудиозаписи с усовершенствованной оценкой меры близости

Предложен метод быстрого поиска неизвестного аудиофрагмента в корпусе аудиозаписей с усовершенствованной оценкой степени близости, позволяющий повысить достоверность результатов поиска. Вследствие применения усовершенствованной оценки степени близости между неизвестным аудиофрагментом и аудиозаписями, содержащимися в корпусе, существенно увеличен уровень различимости собственной аудиозаписи.

Запропоновано метод швидкого пошуку невідомого аудіофрагмента в корпусі аудіозаписів з удосконаленою оцінкою міри близькості, що дозволяє підвищити достовірність результатів пошуку. В результаті застосування вдосконаленої оцінки міри близькості між невідомим аудіофрагментом і аудіозаписями, які містяться в корпусі, істотно збільшено рівень розрізнення власного аудіозапису.

Введение. Рост объемов мультимедийной информации, в частности аудиоинформации, которая передается и обрабатывается в компьютерных системах, обусловила необходимость автоматизации процессов анализа и поиска данных. Поэтому в современных системах обработки аудиоинформации возникает необходимость автоматического поиска музыкальных произведений (МП) на основе аудиоконтента в базах данных. С учетом больших объемов аудиоинформации, хранящейся в корпусе МП, существенное значение имеет скорость и надежность поиска. Для повышения надежности результатов поиска в статье предлагается метод быстрого поиска фрагмента аудиозаписи в корпусе МП с усовершенствованной оценкой степени близости между неизвестным аудиофрагментом и шаблонами, позволяющий повысить достоверность принятия решения при поиске путем увеличения степени различимости шаблона собственного МП с остальными аудиозаписями.

Постановка задачи

Цель данной статьи – повышение достоверности принятия решения в процессе поиска шаблона собственного МП (далее в статье – собственного шаблона) для неизвестного аудиофрагмента благодаря применению усовершенствованной оценки степени близости между фрагментом и шаблонами, которая позволяет существенно увеличить уровень различимости шаблона собственного МП с другими.

Для повышения достоверности принятия решения при поиске собственного шаблона для неизвестного фрагмента необходимо решить следующие задачи:

- выбрать параметры, которые позволили бы однозначно и компактно его описать;
- предложить оценку степени близости неизвестного фрагмента МП и предварительно созданных шаблонов аудиозаписей, обеспечивающую высокий уровень различимости шаблона собственного МП и остальных;
- провести экспериментальные исследования.

Результатом поиска будет шаблон МП в корпусе с минимальным различием относительно входной аудиозаписи. Считаем, что аудиозапись, которую надо определить, точно совпадает с одним из шаблонов корпуса.

Выбор математической модели аудиосигнала мел-частотных кепстральных коэффициентов

Схема определения соответствия МП на основе контента базируется на использовании аудиофайла для построения модели аудиосигнала. Сравнить непосредственно звуковые сигналы во временной области неэффективно, поэтому аудиозапись разбивают на фреймы (фрагменты по 10 – 30 мс), для которых характеристики сигнала остаются относительно устойчивыми. Для каждого фрейма выполняется спектральный анализ, на основе которого так или иначе вычисляется значение вектора параметров (параметризация). Много различных параметров предложено в литературе [1, 2]. В данной статье в качестве параметров выбраны мел-частотные кепстральные коэффициенты (*MFC* – *Mel Frequency Cepstral Coefficients*), которые впервые было предложено использовать в системах распознавания речи и диктора [3], в дальнейшем они получили

широкое применение в процессе информационного поиска музыки (MIR) [4, 5]. В результате параметризации получаем описание МП в виде файла, содержащего вектора MFCC.

Таким образом, если частота дискретизации звукового файла – 44,1 кГц, длина фрейма – 20 мс с перекрытием 0,5 фрейма, то среднестатистическое МП продолжительностью три мин. характеризуют примерно восемь млн. отсчетов, или, с учетом перекрытия фреймов, приблизительно 18 тыс. фреймов, каждый из которых описывается вектором параметров MFCC размерности 13 (рис. 1).

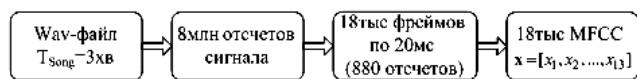


Рис. 1. Описание wav-файла (3 мин.)

Как видно, параметризация позволяет уменьшить количество информации, необходимой для описания МП, в десятки раз:

$$\frac{8\,000\,000 \text{ отсчетов}}{18\,000 \text{ MFCC} \cdot 13} = \frac{8\,000\,000}{234\,000} \approx 34 \text{ раза.}$$

Таким образом, коэффициенты MFCC есть компактным представлением спектральной огибающей, что при распознавании МП позволяет успешно заменить миллионы отсчетов аудио-файла.

Сравнение неизвестного музыкального произведения с шаблонами по приведенному расстоянию

После выбора параметров для описания МП, необходимо перейти к сравнению неизвестного МП с аудиозаписями корпуса и определению шаблона, различие с которым будет минимальным. Для того чтобы идентифицировать неизвестную аудиозапись, необходимо иметь критерий сравнения. Как правило, таким критерием есть расстояние D . Сравнение по выбранным параметрам должно обеспечить высокий уровень различимости собственного шаблона \tilde{X} и шаблонов других произведений \tilde{Y} , т.е. в результате сравнения неизвестного МП X с собственным шаблоном погрешность (по расстоянию) между ними должна быть минимальной, и как можно большей – для шаблонов других произведений:

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

где \tilde{X} – множество векторов параметров собственного шаблона; \tilde{Y} – множество векторов параметров шаблона иного МП.

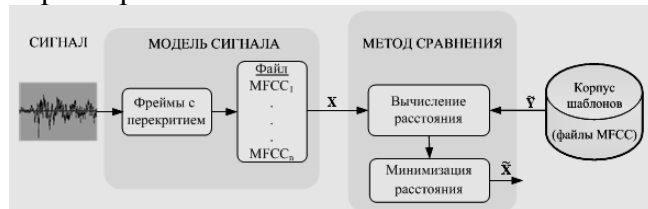


Рис. 2. Общая схема поиска МП

Самым простым и очевидным подходом для определения близости между наборами параметров MFCC неизвестного МП и аудиозаписей в корпусе есть сравнение MFCC на основе наиболее распространенной евклидовой метрики (EM), точнее квадрата евклидова расстояния. Формула невзвешенного евклидова расстояния между вектором параметров $x = (x_1, x_2, \dots, x_d)$, МП, которое нужно идентифицировать, и вектором шаблона $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_d)$:

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

Соответственно расстояние (погрешность) между файлами параметров неизвестного произведения и шаблона можно найти как сумму расстояний:

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

В идеальном случае при таком подходе расстояние между файлами параметров MFCC одного и того же МП будет равна нулю, для разных – отличной от нуля:

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

Отметим, что даже для одного и того же МП аудиозаписи могут отличаться, например: в начале записи может идти тишина, мелодия другого МП и т.д.; записи могут иметь разный темп, продолжительность. Это означает, что при смещении фреймов во времени расстояние к собственному шаблону $D(X, \tilde{X}) \neq 0$, т.е. условие различимости собственного МП (4) не выполняется, следовательно, непосредствен-

ное сравнение файлов параметров по евклидову расстоянию не подходит для задачи поиска аудиофрагмента в корпусе аудиозаписей.

Очевидно, что в большинстве случаев МП характеризуется определенной периодичностью, что заключается в наличии идентичных или очень похожих по тексту и характеру мелодии фрагментов. Соответственно можно говорить об избыточности данных, которыми описывается аудиозапись, и сократить количество параметров для ее описания. Учитывая это целесообразно применить методы кластерного анализа для формирования корпуса аудиозаписей, приведенные далее. Использование кластеризации для формирования эталонов корпуса позволит уменьшить объемы памяти, необходимые для их хранения.

Основные этапы сравнения файлов параметров неизвестного МП с определенным шаблоном в корпусе:

- Поиск минимального евклидова расстояния $D \min^2$ между текущим вектором параметров $\mathbf{x} = (x_1, x_2, \dots, x_d)$ из множества параметров $\mathbf{X} = \{\mathbf{x}\}$, $|\mathbf{X}| = n$ МП, которое нужно идентифицировать, и множеством векторов-центроидов $\tilde{\mathbf{Y}} = \{\tilde{\mathbf{y}}\}$, $|\tilde{\mathbf{Y}}| = m$, $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_d)$ шаблона:

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

- Вычисление оценки расстояния в целом к шаблону в виде суммы квадратов минимальных расстояний $D \min^2$:

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

Вместе с тем, поскольку МП имеют разную продолжительность, каждое произведение характеризуется собственным количеством фреймов, представленных векторами *MFCC*. При кластеризации вычисляется одинаковое количество центроидов для всех шаблонов. Это приводит к тому, что разные произведения находятся в неравных условиях, т.е. для произведений, длительность которых больше, начальная погрешность (между файлами того же произведения до и после кластеризации) тоже будет

больше, так как в этом случае на каждый кластер будет приходиться больше векторов параметров. Избежать этого можно путем деления расстояния до шаблона, определенного в формуле (6), на количество фреймов МП n . Назовем эту величину приведенным расстоянием ($ПР - D_{ПВ}$) аудиозаписи:

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

Как видно из формулы (7) характеристика $D_{ПВ}$ не зависит от количества фреймов (продолжительности записи). Таким образом, ее можно использовать в качестве критерия принятия решения как для записи в целом, так и для ее отдельного аудиофрагмента. Обоснование этого приведено в [6], там же определена длительность фрагмента МП в 5с (500 фреймов или 500 векторов параметров *MFCC*) достаточная для поиска аудиозаписи.

Создание корпуса шаблонов музыкального произведения на основе кластеризации

Методы кластеризации широко используются в системах распознавания аудиосигнала [7]. Задачу кластеризации можно сформулировать так: заданный набор из n векторов размерности d необходимо разбить на подмножества в соответствии с критерием минимизации искажения $e_i^2 \rightarrow \min$. Существуют различные пути оценивания искажения, но в большинстве прикладных реализаций используют сумму среднеквадратических евклидовых расстояний между центром кластера (центроидом) \mathbf{c}_i и векторами параметров, которые к нему относятся $\mathbf{X}_i = \{\mathbf{x}\}$, $\mathbf{X}_i \subset \mathbf{X}$ [8, 9], т.е.:

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

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

где N_i – количество векторов, принадлежащих центроиду \mathbf{c}_i .

Таким образом, шаблоны МП можно описать центроидами кластеров параметров *MFCC*.

Поскольку между каждым центроидом \mathbf{c}_i и векторами, относящимися к нему, есть погрешность e_i^2 , то расстояние между файлами параметров, описывающих одно и то же МП до кластеризации и после (даже если аудиозаписи были идентичными), будет положительным и равным величине суммарной погрешности кластеризации E^2 .

Итак, сформированный корпус шаблонов МП для идентификации неизвестного фрагмента аудиозаписи фактически представляет собой файл, содержащий центроиды *MFCC* всех шаблонов МП.

Применение *kd*-деревьев для поиска фрагмента музыкального произведения в корпусе шаблонов

Самое простое решение общеизвестной задачи поиска ближайшего соседа, а в нашем случае – ближайшего центроида в корпусе аудиозаписей – это вычисление расстояния между вектором неизвестного аудиофрагмента, который нужно идентифицировать, и всеми векторами-центроидами шаблонов. Время выполнения полного поиска пропорционально $O(dN)$, где N – количество центроидов в корпусе, d – размерность векторов. Однако применение такого подхода для поиска в больших базах данных практически нереализуемо. Поэтому для уменьшения сложности вычислений при поиске фрагмента аудиозаписи в сформированном корпусе был применен простой способ разбиения пространства – k -мерное дерево (*kd*-дерево). Центроиды аудиозаписей в корпусе были упорядочены на основе *kd*-дерева [10]. Упорядочение выполняется очень быстро, поскольку разбиение пространства происходит только вдоль осей данных, соответственно при этом не нужно вычислять d -мерные расстояния. В работе [11] предложено усовершенствованную процедуру поиска по *kd*-дереву, примененную для поиска ближайшего центроида в корпусе шаблонов в процессе поиска фрагмента МП.

Чтобы обеспечить нахождение ближайшего вектора, поиск, кроме прямой фазы (спуска по дереву), должен иметь также обратную (поиск элементов дерева в заданном диапазоне). Во время прямого поиска фиксируются все расстояния к узлам d_i . Прямая фаза завершается вычислени-

ем расстояния $D_{\min} = D_k$ к соответствующей терминальной вершине, в которой лежат неизвестный вектор, а также потенциально ближайший элемент дерева (центроид шаблона в корпусе), что задает на обратной фазе радиус поиска. В свою очередь на обратной фазе поиска вычисляются расстояния d_i лишь к тем узлам дерева, которые могут обеспечить выполнение $d_i < D_{\min}$. Если для соответствующего листа выполняется условие $d_k < D_{\min}$, радиус поиска корректируют $D_{\min} = D_k$ [11, 12].

Рассмотренная процедура поиска гарантирует нахождение ближайшего вектора по формуле (1), однако нуждается в большем количестве измерений расстояния, чем $\log_2 n$.

Отметим, что поиск на основе *kd*-дерева – очень эффективен только для небольших размерностей ($d < 20$). Для $d < 20$ среднее время поиска ближайшего вектора в корпусе, упорядоченного на основе *kd*-дерева, в среднем растет пропорционально $O(h) = O(d \cdot \log(N+1))$, где h – высота дерева [11]. Для больших размерностей вычислительная сложность может достигать $O(d \cdot N)$.

Приведенная процедура поиска обеспечивает получение не одного ближайшего вектора, а некоторого множества векторов, упорядоченных по возрастанию расстояния согласно формуле (2). Это возможно, если данные о расстоянии к уже пройденным терминальным вершинам сохраняются, и одновременно осуществляется их упорядочение по возрастанию расстояния к неизвестному вектору. Благодаря этому дополнительное нахождение нескольких ближайших векторов не требует много времени.

Оценки степени близости неизвестного фрагмента с шаблонами музыкального произведения

Для каждого фрейма (вектора) \mathbf{x}_l неизвестного аудиофрагмента сначала выполняется быстрый поиск в упорядоченном на основе *kd*-дерева корпусе шаблонов МП, в процессе которого по ЕМ по формуле (2) отбирается множество \mathbf{C} векторов-центроидов шаблонов (кандидатов на ближайший вектор), упорядоченных по возрастанию расстояния:

$$\mathbf{C} \subset \mathbf{T}, \mathbf{C} = \{\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2, \dots, \tilde{\mathbf{Y}}_k\},$$

$$|\tilde{\mathbf{Y}}| = m, |\mathbf{C}| = k, |\mathbf{T}| = N = m \cdot Q, k \leq N,$$

где k – количество векторов шаблонов в списке ближайших к фрейму неизвестного фрагмента; m – количество центроидов (векторов) шаблона; Q – общее количество аудиозаписей в корпусе; N – общее количество центроидов в корпусе.

Для каждого из k ближайших центроидов множества \mathbf{C} хранятся расстояние к фрейму неизвестного фрагмента $D_p^2(\mathbf{x}, \tilde{\mathbf{y}}_{qj})$ и индекс соответствующего шаблона МП i_{pq} в корпусе:

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

После этого определяется шаблон, ближайший к неизвестному аудиофрагменту.

В табл. 1 показаны результаты для первых десяти из списка k ближайших центроидов шаблонов для первого и 145-го фреймов неизвестного МП, полученные поиском на основе kd -дерева. Индекс собственного шаблона (ближайшего суммарно по всем фреймам фрагмента) выделен цветом.

Таблица 1

№ ПП	Индексы ближайших шаблонов (и расстояния) для первого и 145-го фреймов			
	1		145	
1	397	(0,288)	2	(0,431)
2	694	(0,306)	553	(1,115)
3	220	(0,319)	457	(1,159)
4	17	(0,358)	479	(1,177)
5	19	(0,359)	407	(1,238)
6	2	(0,380)	2	(1,269)
7	453	(0,385)	19	(1,273)
8	409	(0,396)	952	(1,284)
9	988	(0,419)	546	(1,317)
10	659	(0,425)	2	(1,353)

Из результатов табл. 1 видно, что центроиды собственного шаблона могут находиться не на первых позициях (первый фрейм) в списке k ближайших. Также можно наблюдать на примере 145-го фрейма, что центроиды собственной аудиозаписи в списке ближайших могут встречаться несколько раз. Кроме того, полученные данные еще раз подтверждают то, что фрагмент для поиска лучше брать не с 7-го начала МП.

На основе представленных в табл. 1 данных и их анализа рассмотрим возможные варианты

вычисления степени близости неизвестного фрагмента с шаблонами в корпусе МП.

Вычисление степени близости по приведенному расстоянию

• Для текущего фрейма \mathbf{x}_l неизвестного фрагмента расстояние к каждому шаблону определяется в соответствии с правилом:

$$Distance_q = \begin{cases} \min(D_p^2), \forall q \in \mathbf{I} \\ D_p^2, p = k, \forall q \notin \mathbf{I}, \end{cases} \quad (10)$$

когда на данном фрейме расстоянием к шаблону будет считаться наименьшее расстояние среди его центроидов, попавших в список ближайших. Расстояние для аудиозаписей в корпусе, центроиды которых не попали в список k ближайших, считается равным расстоянию для последнего k -го центроида из списка ближайших. Последнее фактически означает, что искусственно занижается расстояние к тем шаблонам, которые не попали в список k ближайших, и соответственно уменьшается уровень различимости собственного шаблона и остальных аудиозаписей в корпусе.

• Для каждого шаблона $\tilde{\mathbf{Y}}_q$ вычисляется сумма расстояний по всем n фреймам неизвестного фрагмента.

$$D(\mathbf{X}, \tilde{\mathbf{Y}}_q) = \sum_{l=1}^n Distance_{q_l}. \quad (11)$$

• На последнем шаге находится минимум приведенного расстояния, определенного ранее в формуле (7):

$$D(\mathbf{X}, \tilde{\mathbf{X}}) = \min_q D_{\text{ПВ}q} = \min_q \frac{D(\mathbf{X}, \tilde{\mathbf{Y}}_q)}{n}. \quad (12)$$

Считаем, что неизвестный аудиофрагмент \mathbf{X} принадлежит тому шаблону $\tilde{\mathbf{X}}$ из корпуса МП, который обеспечивает минимум в формуле (12).

Вычисление степени близости по количеству попаданий шаблона в список k ближайших центроидов (Hits)

• Для каждого фрейма \mathbf{x}_l неизвестного фрагмента фиксируется количество попаданий центроидов q -го шаблона в список k ближайших:

$$Hits_q = \sum_{p=1}^k h_p, \quad \text{где } h_p = \begin{cases} 1, \forall q \in \mathbf{I} \\ 0, \forall q \notin \mathbf{I} \end{cases}. \quad (13)$$

- Для каждого шаблона \tilde{Y}_q находится суммарное значение количества попаданий по всем n фреймам неизвестного фрагмента.

- Собственным шаблоном МП считается тот, для которого значение оценки меры близости согласно формуле (13) по всем n фреймам, есть самым большим:

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

При таком подходе разница между собственной аудиозаписью и другими растет, при этом сложность вычислений уменьшается.

Вычисление степени близости по взвешенному количеству попаданий шаблона в список k ближайших центроидов (Weighted Hits)

При этом подходе к оценке степени близости между неизвестным фрагментом и шаблонами учитываются преимущества двух предыдущих, т.е. учитываются как расстояния D_p^2 , так и количество попаданий h_p центроидов шаблона в список k ближайших, а также позиция шаблона в списке ближайших, что будет отражено в формуле (15).

- Для каждого фрейма x_l фрагмента вычисляется величина, названная взвешенным количеством попаданий центроидов q -го шаблона в список k ближайших по формуле:

$$\mathbf{Weighted Hits}_q = \sum_{p=1}^k (D_k^2 - D_p^2) \cdot h_p. \quad (15)$$

- Аналогичным образом (как и в предыдущем подходе) для каждой аудиозаписи \tilde{Y}_q в корпусе находится суммарное значение взвешенного количества попаданий по всем n фреймам неизвестного фрагмента.

- Собственным шаблоном МП считается тот, для которого значение оценки меры близости согласно формуле (15) по всем фреймам, разделенное на общее их количество (как в первом подразделе), является наибольшим (как и во втором):

$$WH(\mathbf{X}, \tilde{\mathbf{X}}) = \max_q \sum_{l=1}^n \frac{\mathbf{Weighted Hits}_q}{n}. \quad (16)$$

Этот усовершенствованный подход оценивания степени близости между неизвестным

фрагментом и шаблонами должен позволить уравновесить преимущества и недостатки предыдущих двух.

Экспериментальные результаты

Для проведения экспериментальных исследований был сформирован корпус из 1000 аудиозаписей МП. Все МП имели формат *wav* (*mono*) с частотой дискретизации 44,1 кГц. Предварительно из аудиофайлов удалена *тишина* в начале и конце записей. В процессе формирования корпуса МП аудиозаписи шаблонов были разбиты на фреймы по 20 мс с перекрытием 10 мс и для каждого фрейма вычислен вектор параметров *MFCC* размерности 13. Последовательности векторов параметров *MFCC*, описывающих МП, были кластеризованы, используя усовершенствованный метод кластеризации k -средних, предложенный в [13]. После кластеризации каждый шаблон в корпусе был представлен 1000 центроидов *MFCC* (в среднем около 20 векторов на кластер). Фрагменты продолжительностью 5с для идентификации выбирались случайным образом из эталонов сформированного корпуса МП. Применение поиска по *kd*-дереву и оценки меры близости к шаблонам согласно выражениям (10) – (12) в сравнении с полным поиском приводит к некоторому уменьшению степени различимости собственной аудиозаписи и остальных шаблонов корпуса (рис. 3), однако разница остается достаточной для принятия правильного решения.

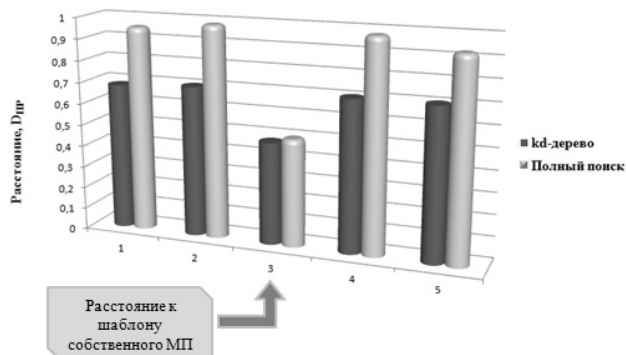


Рис. 3. Сравнение результатов полного поиска и поиска на основе *kd*-дерева

На рис. 4 показаны результаты сравнения неизвестного фрагмента с шаблонами (собственный шаблон под № 3) по предложенным в разделе формулам (11) – (16) для оценки степени

их близости, и сводную гистограмму, на которой представлены все варианты оценивания степени близости шаблонов и неизвестного фрагмента в относительных единицах для достижения единого масштаба.

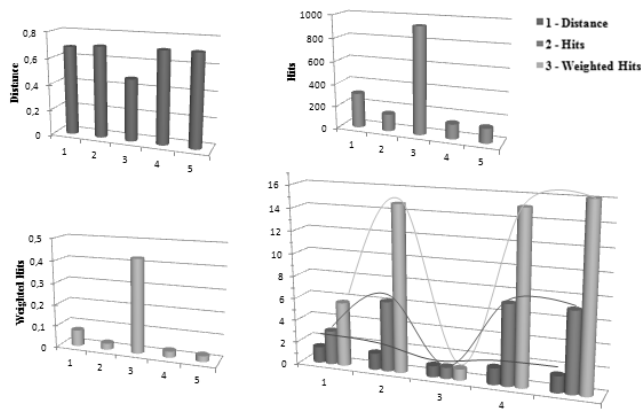


Рис. 4. Сравнение оценок степени близости неизвестного фрагмента с шаблонами

Как видно из приведенных результатов, оценивание степени близости по взвешенному количеству попаданий шаблона в список k ближайших (*Weighted Hits*) обеспечивает высокий уровень различимости собственного и чужих шаблонов.

Также на уровень различимости собственного шаблона и остальных влияет количество центроидов, которыми представлены аудиозаписи в корпусе, продолжительность фрагмента для поиска и количество ближайших центроидов, получаемых в результате поиска по kd -дереву. Последний вариант наиболее приемлем для повышения уровня различимости, так как не приводит к значительному росту времени поиска. Рассматривались результаты для 20 и 50 ближайших кандидатов. В табл. 2 для пяти фрагментов представлены результаты, полученные при выборе собственной аудиозаписи в корпусе по оценкам степени близости неизвестного фрагмента с шаблонами, описанным ранее. Для каждой из оценок степени близости показаны значения, полученные для собственного шаблона (выделены цветом), и наиболее близкого среди других.

Полученные данные показывают, что, с увеличением количества ближайших соседей до 50, разница в оценках возрастает, что увеличивает

степень различимости между собственным шаблоном и остальными.

Т а б л и ц а 2

k	20					50				
№ фрагм.	1	2	3	4	5	1	2	3	4	5
$D_{ПВ}$	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

Оценка расстояния к собственной аудиозаписи, полученная по формуле (16), при $k = 20$ в среднем в восемь раз превышает аналогичную оценку расстояния к другим шаблонам. Таким образом, усовершенствованная оценка близости значительно повышает достоверность принятия решения в процессе поиска шаблона собственного МП и обеспечивает возможность для дальнейшего расширения корпуса МП.

Отметим, что для выполнения поиска в реальных архивах аудиозаписей, что содержат сотни тысяч, миллионы аудиозаписей, критическим требованием есть высокая скорость поиска. Выполнение не точного, а приближенного поиска векторов (когда найденный вектор не обязательно ближайший, но достаточно близкий) на основе kd -дерева позволяет достичь значительной экономии времени, однако приводит к снижению достоверности результатов поиска. Поэтому для уменьшения сложности вычислений при сохранении достоверности результатов целесообразно для больших архивов аудиозаписей выполнять комбинированный поиск, сочетающий быстрый *неточный* поиск с применением kd -дерева нескольких ближайших аудиозаписей корпуса для заданного аудиофрагмента на первом этапе поиска, среди которых на втором этапе происходит уточнение собственной аудиозаписи с помощью полного перебора.

Получил дальнейшее развитие метод поиска на основе kd -дерева с усовершенствованной оценкой меры близости по взвешенному количеству попаданий в список ближайших центроидов. Предложенный метод позволил увеличить полноту и релевантность результатов поиска, а именно: для аудиофрагмента длительностью 5 с полнота поиска составляет 99,8 про-

центров, релевантность – 97,4 процентов, для аудиофрагмента 1 с – 99,6 и 94,3 процентов соответственно, что превышает аналогичные показатели существующих методов.

Заключение. В статье предложен метод быстрого поиска аудиофрагментов продолжительностью 5с на основе *kd*-дерева с усовершенствованной оценкой степени близости между неизвестным фрагментом и аудиозаписями в корпусе, позволяющий повысить достоверность результатов поиска. Предложены аналитические соотношения для оценки степени близости неизвестного фрагмента с шаблонами, которые базируются на вычислении расстояния от фрагмента к заданному числу ближайших центроидов и количества попаданий в список ближайших соседей. Применение усовершенствованной оценки степени близости (по взвешенному количеству попаданий в список *k* ближайших центроидов) между неизвестным аудиофрагментом и шаблонами существенно увеличивает уровень различимости собственной аудиозаписи в корпусе. Оценка расстояния к собственной аудиозаписи при $k = 20$ в среднем в восемь раз превышает аналогичную оценку расстояния к другим аудиозаписям корпуса.

1. Wang Y., Liu Z., Huang J.C. Multimedia content analysis using both audio and visual cues // IEEE signal processing magazine. – 2000. – N 17. – P. 12–36.
2. Grosche P., Müller M., Serrà J. Audio content-based music retrieval // Dagstuhl Follow-Ups Multimodal Music Processing. – Dagstuhl, Germany. – 2012. – 3. – P. 157–175.
3. Ganchev T., Fakotakis N., Kokkinakis G. Comparative evaluation of various mfcc implementations on the

- speaker verification task // Proc. of 9th Int. Conf. on Speech and Comp., SPECOM'05. – 2005. – P. 191–194.
4. Logan B., Salomon A. A music similarity function based on signal analysis // Proc. IEEE Int. Conf. Multimedia Expo. – 2001. – P. 745–748.
5. Tzanetakis G., Cook P. Musical genre classification of audio signals // IEEE Trans. Speech Audio Process. – 2002. – 10, N 5. – P. 293–301.
6. Ткаченко О.М., Грійо Тукало О.Ф. Підхід до оцінювання тривалості фрагмента для пошуку музичного твору за заданим шаблоном // Інформаційні технології та комп'ютерна інженерія. – Вінниця: ВНТУ, 2014. – № 1. – С. 31–40.
7. Сажок Н.Н. Кластеризация слов при построении лингвистической модели для автоматического распознавания речевого сигнала // Кибернетика и вычислительная техника. – 2012. – № 4. – С. 59–66.
8. Gersho A., Gray R.M. Vector Quantization and Signal Compression. – Boston: Kluwer Academic, 1992. – 760 p.
9. Jain A.K., Dubes R.C. Algorithms for Clustering Data. – Englewood Cliffs, N.J.: Prentice Hall. – 1988. – 334 p.
10. Moore A. An introductory tutorial on KD trees // Technical Report. – 1991. – N 209 (Extract from PhD Thesis). – Computer Laboratory, University of Cambridge. – <http://www.autonlab.org/autonweb/14665/version/2/paet/5/data/moore-tutorial.pdf>
11. Friedman J.H., Bentley J.L., Finkel R.A. An algorithm for finding best matches in logarithmic expected time // ACM Transactions on Mathematical Software, 1977. – 3(3). – P. 209–226.
12. Arya S., Mount D.M. Algorithms for fast vector quantization // Proc. of DCC '93: Data Compression Conference. – 1993. – P. 381–390.
13. Метод кластеризації на основі послідовного запуску *k*-середніх з удосконаленням вибором кандидата на нову позицію вставки / О.М. Ткаченко, О.Ф. Грійо Тукало, О.В. Дзись та ін. // Наукові праці ВНТУ. – 2012. – № 2. – 10 с. – <http://praci.vntu.edu.ua/article/view/2325>

E-mail: alextk1960@gmail.com, xxmargox@gmail.com
© А.Н. Ткаченко, О.Ф. Грійо Тукало, 2015

UDC 621.39

A.N. Tkachenko, O.F. Griyo Tukalo

The Method of Audio Fragment Fast Search with an Improved Estimation of the Measure Proximity

The development of the Internet, digital media, multimedia technologies and new ways of processing audio information generates enormous interest and attention to the ways in which information technology can be applied. With access to the huge archives of multimedia, virtually anywhere and at any time, it is necessary to propose the new ways to navigate and interact with these large audio archives. This paper proposes a method for the fast search of the unknown audio fragment in the corpus of audio recordings with the improved estimation of the measure proximity, thus enhancing the reliability of the audio search. Through the use of the improved estimation of the measure proximity between the unknown fragment and the audio recordings contained in the corpus, the level of distinguishability of its own audio recording is significantly increased. Experimental results show that the overall distance to its own audio on average 8 times higher than estimate distances to other audio recordings in the corpus.