Д. Ю. Мухин

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ВАРЬИРОВАНИЯ ПРИМЕНИТЕЛЬНО К MIDI-ФАЙЛАМ*

ВВЕДЕНИЕ

Одним из существенных элементов процесса творчества является процедура транспонирования (переноса) некоторой структуры из одной ситуации в другую. Мы займемся исследованием закономерностей таких переносов и их места в общем процессе творчества, также изучением методов формирования возникающих при этом вариантов и способов их оценки.

Известно, что восприятие характеризуется такой качественной особенностью, как структурность. Структура, целостная и неразложенная на составляющие части, носитель некоторого образа. Возможность перенесения структуры из одной ситуации в другую, в новые условия, или транспонируемость, является основным свойством структуры.

В нашем случае структурой будет являться мелодия, формализованная в виде символьной последовательности. Каждой мелодии мы поставим в соответствие три числовых последовательности.

1. Последовательность номеров нот. Все ноты, начиная от ноты «до» первой октавы, мы пронумеровываем натуральными числами от 0 до 128.

2. Последовательность интервалов. Последовательность звуковысотных интервалов между соседними нотами. Если каждый интервал выразить числом полутонов, то такую последовательность можно записать в виде цепочки целых чисел (знак «-» будет показывать движение текущей ноты вниз относительно непосредственно предшествующей ей ноты, отсутствие знака перед интервалом — движение текущей ноты вверх).

3. **Последовательность длительностей.**

Одним из методов изучения мышления является варьирование ситуаций. При изменении некоторой первоначальной ситуации получается варьиро-

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 01-01-794) и Министерства образования РФ.

ванная ситуация, но при восприятии исходной и варьированной ситуаций обнаруживается их известная общность.

Наряду с новым в варьированной ситуации ощущается непосредственная связь с первоначальной ситуацией, элемент повторяемости или постоянства. Это обстоятельство связано с наличием инвариантов преобразования — таких элементов ситуации, которые остаются неизменными, постоянными при любой трансформации. Иначе говоря, при варьировании происходит перенос определенных отношений элементов из одной ситуации в другую, т.е. сохранение некоторых исходных структурных свойств.

Наличие инвариантов в варьированной и исходной ситуациях часто маскируется другими элементами, которые как бы меняют лицо первоначальной ситуации. Это трансформанты — элементы, изменяющиеся при преобразовании. Маскирующие элементы часто до неузнаваемости трансформируют ситуацию и весьма затрудняют обнаружение инвариантов преобразования.

Все трансформации можно разделить на три основных класса.

1. Трансформации звуковысотности. При таких трансформациях осуществляются изменения в последовательности высот звуков. Например, элементарной трансформацией такого типа будет перенос всей мелодии по вертикали. В рамках нашей модели — увеличение каждого из чисел (которыми закодирована каждая нота нашей мелодии) на фиксированное число n.

2. Трансформации ритма. При таких трансформациях изменения вносятся в ритмическую часть мелодии. Например, изменение длительностей нот или метрических акцентов. В рамках нашей модели: изменяться будут числа, которыми закодированы длительности нот.

3. Орнаментальные трансформации. При трансформациях такого типа изменения вносятся как в звуковысотную, так и в ритмическую часть мелодии. Примером будет являться такая комбинация трансформаций: каждая длительность ноты дробится на несколько частей (одну, две, три, четыре), в каждом дроблении высота первой ноты совпадает с высотой ноты в теме до дробления.

При прослушивании примеров к каждому классу трансформаций, мы заметили, что в некоторых образах искомая мелодия угадывалась сразу, а в некоторых была изменена до неузнаваемости. Секрет в том, что эти разные последовательности звуков объединяет какое-то одно из их общих качеств.

В примере для первого класса трансформаций была одна и та же последовательность звуковысотных интервалов между соседними нотами. Здесь именно эта последовательность интервалов и способствует восприятию одной и той же мелодии при прослушивании разных последовательностей нот. Однако та инвариантная структура, которая служит носителем образа мелодии, не сводится к одной лишь неизменной последовательности интервалов. Такая структура в нашем первом примере состоит из двух инвариантов — последовательности интервалов и ритма. Если же разложить ее на составляющие части — ритм и последовательность интервалов, то взятая в отдельности каждая часть, являясь инвариантом, тем не менее, может и не быть носителем первоначального образа.

В примерах для второго класса трансформаций в первом случае инвариантами являлись последовательность высот звуков и ритм, а во втором — только ритм.

В примерах для третьего класса трансформаций инвариантом была неизменная последовательность высот первых нот каждой четверки с сохранением метрических акцентов.

Итак, тот или другой инвариант обеспечивает сохранение первоначального образа — образа исходной ситуации при варьировании. Для этого необходима инвариантная структура — совокупность некоторых инвариантов, которая и является носителем данного образа.

Таким образом, инвариантная структура определяется не элементами, ее образующими, а системой отношений между этими элементами. При транспонировании осуществляется изменение элементов, звуков, составляющих мелодию, по высоте и другим параметрам, но сохраняется система отношений между ними.

1. ПОСТАНОВКА ЗАДАЧИ

На входе имеется несколько MIDI-файлов, на первой дорожке каждого из которых записана одноголосая мелодия. Один из файлов — эталон (исходная мелодия), остальные — эталон, трансформированный разными способами. Наша задача заключается в том, чтобы установить степень сходства между трансформированными мелодиями и мелодией-эталоном и проанализировать изменение степени в зависимости от произведенных трансформаций.

Далее, получаем результаты для MIDI-файлов, каждый из которых конвертируется в WAVE-файл и проводим уже спектральный анализ (основным объектом изучения уже будет звуковой сигнал).

Результаты обоих типов сравниваются.

2. ЗАДАЧА НЕЧЕТКОГО СОПОСТАВЛЕНИЯ СТРОК

Строка длины $|x|=m$ записывается как $x_1x_2...x_m$, где x_i представляет i-й символ x. Подстрока $x_ix_{i+1}...x_j$ строки x, где $i \le j \le m$, будет обозначаться $x(i,j)$.

Обобщенная задача сопоставления строк, включающая в себя нахождение подстрок строки текста, близких к заданному образцу строки, называется также задачей нечеткого сопоставления строк. Задачу нечеткого сопоставления строк можно сформулировать следующим образом.

Пусть даны образец x, $|x|=m$, и текст y, $|y|=n$. Пусть даны также целое $k>0$ и функция расстояния d. Требуется найти все подстроки s текста y такие, что $d(x,s)<k$. Здесь и далее d — метрика.

Расстояние Хемминга [Hamming, 1982] между двумя строками одинаковой длины определяется как число позиций, в которых символы не совпадают. Это эквивалентно минимальной цене преобразования первой строки во вторую в случае, когда разрешена только операция замены с единичным весом. Если допускается сравнение строк разной длины, то как правило, требуются также вставка и удаление. Если придать им тот же вес, что и замене, минимальная общая цена преобразования будет равна одной из метрик, предложенных Левенштейном [Levenstein, 1965]. Нас будет интересовать метрика Левенштейна, т.к. в случае трансформации ритмической части мелодии длины наших числовых последовательностей могут не совпадать.

Задача, таким образом, состоит в том, чтобы при заданной функции расстояния найти все подстроки текста, отстоящие от образца не более чем на k. Если d является расстоянием Хемминга, задача называется сопоставлением строк с k несовпадениями, если же d — расстояние Левенштейна, задача называется сопоставлением строк с различиями (или иногда ошибками).

2.1. Алгоритм Вагнера—Фишера

В методе динамического программирования последовательно по предыдущим значениям вычисляются расстояния между все более и более длинными префиксами двух строк — до получения окончательного результата. Опишем этот процесс более подробно.

Пусть $d_{i,j}$ есть расстояние между префиксами строк x и y, длины которых равны соответственно i и j, т. е. $d_{i,j} = d(x(1,i), y(1,j))$.

Цену преобразования символа a в b обозначим через $w(a,b)$. Таким образом, $w(a,b)$ — это цена замены одного символа на другой, когда $a \neq b$, $w(a,\varepsilon)$ — цена удаления a, а $w(\varepsilon,b)$ — цена вставки b. Заметим, что в случае, когда выполнены нижеследующие условия, d является расстоянием Левенштейна

$$w(a,\varepsilon) = 1,$$
$$w(\varepsilon,b) = 1,$$
$$w(a,b) = 1, \text{ если } a \neq b,$$
$$w(a,b) = 1, \text{ если } a = b.$$

В процессе вычислений значения $d_{i,j}$ записываются в массив $(m+1)\times(n+1)$, а вычисляются они с помощью следующего рекуррентного соотношения:

$$d_{i,j} = \min\{d_{i-1,j} + w(x_i,\varepsilon), d_{i,j-1} + w(\varepsilon, y_j), d_{i-1,j-1} + w(x_i, y_i)\}.$$

Оно выводится следующим образом. Если предположить, что известна цена преобразования $x(1, i-1)$ в $y(1, j)$, то цену преобразования $x(1,i)$ в $y(1, j)$ мы получим, добавив к ней цену удаления x_i. Аналогично, цену преобразования $x(1,i)$ в $y(1, j)$ можно получить, прибавив цену вставки y_j к цене преобразования $x(1,i)$ в $y(1, j-1)$. Наконец, зная цену преобразования $x(1, i-1)$ в $y(1, j-1)$, цену преобразования $x(1, i)$ в $y(1, j)$ мы получим, прибавив к ней цену замены x_i на y_j. Вспомним, что расстояние $d_{i,j}$ является минимальной ценой преобразования $x(1, i)$ в $y(1, j)$, поэтому из трех указанных выше операций надо выбрать самую дешевую.

Перед тем как начать вычислять $d_{i,j}$, надо установить граничные значения массива. Что касается первого столбца массива, то значение $d_{i,0}$ равно сумме цен удаления первых i символов x. Аналогично, значения $d_{0,j}$ первой строки задаются суммой цен вставки первых j символов y. Итак, имеем следующее:

$$d_{0,0} = 0\,,$$

$$d_{i,0} = \sum_{k=1}^{i} w(x_i, \varepsilon) \text{ для } 1 \le i \le m\,,$$

$$d_{0,j} = \sum_{k=1}^{j} w(\varepsilon, j_k) \text{ для } 1 \le j \le n\,.$$

Для расстояния Левенштейна $d_{i,0} = i$ и $d_{0,j} = j$. Ниже приведен массив, полученный при вычислении расстояния Левенштейна, например, между строками «12345» и «12354». Из него видно, что расстояние между этими строками, т.е. $d_{5,5}$, равно 2:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 4 & 3 & 2 & 1 & 1 & 1 \\ 5 & 4 & 3 & 2 & 1 & 2 \end{pmatrix}.$$

Алгоритм вычисления массива расстояний разработан Вагнером и Фишером (Wagner, Fisher). Можно видеть, что стадия инициализации границ включает $1+m+n$, а основной цикл повторяется mn раз. Таким образом, временная сложность этого алгоритма есть $O(mn)$. Такое положение вещей нас устраивает, так как мы имеем дело с числовыми последовательностями сравнительно малой длины.

2.2. Анализ результатов (MIDI)

Эталон (обозначим его 0-й мелодией) — MIDI-файл с одноголосой мелодией из известного классического произведения. В таблице приведены четыре трансформации звуковысотности — мелодия была сыграна в четырех различных гаммах (назовем трансформированные мелодии соответственно 1-й, 2-й, 3-й и 4-й). Заранее была субъективно установлена степень сходства между эталоном и трансформированными мелодиями.

При сравнении эталона и 1-ой мелодии на слух образ сохранился полностью, при сравнении эталона и последующих мелодий субъективно при прослушивании сходство распределилось по убывающей.

Далее приводятся результаты сравнения с помощью метрики Левенштейна (L):

Сравниваемые мелодии	L для посл-ти нот	L для посл-ти интервалов	L для посл-ти длительностей
0 & 1	24	0	0
0 & 2	5	10	0
0 & 3	10	16	0
0 & 4	13	20	0

В первом случае была применена трансформация звуковысотности — перенос всей мелодии по вертикали на пять нот вверх. Несмотря на то что дистанция для последовательности нот принимает наибольшее значение из всех, образ сохранился полностью. Это произошло из-за наличия инварианта — последовательности интервалов, на что указывает нулевая дистанция для последовательности интервалов.

В остальных трех случаях производились более сложные трансформации звуковысотности, для этого использовалась программа *Ntonyx Style Morpher*. В результате получилось, что меньшему изменению образа соответствуют меньшие значения метрики Левенштейна (L) для последовательностей нот, для последовательностей интервалов и для последовательностей длительностей.

3. ПРЕОБРАЗОВАНИЕ АДАМАРА

Теперь, сконвертировав проанализированные MIDI-файлы в формат WAVE, подойдем к задаче исследования алгоритмов варьирования с точки зрения спектрального анализа звуковых сигналов.

Существует бесконечное количество ортонормированных базисов, по которым можно раскладывать сигналы. Сейчас самым распространённым и наиболее естественным является базис Фурье тригонометрических функций (sin и cos). Но в некоторых ситуациях другие базисы оказываются более оптимальными с точки зрения скорости выполнения преобразования на компьютере, вида коэффициентов и т.д. Может оказаться, что для данного вида сигналов больше подходит преобразование, отличное от Фурье. Поэтому при обработке сигналов желательно иметь достаточно широкий набор различных преобразований для более глубокого анализа. Одним из таких преобразований и является преобразование Уолша-Адамара. Его отличительной особенностью является то, что сигнал раскладывается не по синусам и косинусам, а по функциям Уолша, принимающим значения ±1, что значительно облегчает вычисления на компьютере и позволяет быстро оперировать большими объёмами информации. Для цифровой обработки звуковых сигналов необходимо представить их в цифровом виде, т.е. провести аналого-цифровое преобразование, которое заключается в дискретизации (разбиения области определения) и квантовании (отображении области значений в дискретный диапазон). Амплитуда кодируется целым числом в диапазоне от –32768 до 32767. По теореме Найквиста-Котельникова максимальная частота равна ½ частоты дискретизации. Стандартная дискретизация CD-качества равна 44100Hz, следовательно максимальная воспроизводимая частота — 22050Hz, что не намного превышает максимальную частоту, слышимую человеком. Таким образом, сигнал в цифровом представлении — это последовательность чисел, которую мы и будем преобразовывать с помощью дискретного преобразования Уолша-Адамара.

Рассмотрим матрицу Адамара, которая строится следующим способом:

$$H_4 = \begin{bmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} & \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} & \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \end{bmatrix}, \quad H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_{2^{n-1}} = \begin{bmatrix} H_{2^n} & H_{2^n} \\ H_{2^n} & -H_{2^n} \end{bmatrix}.$$

В этой матрице необходимо упорядочить строки по числу перемен знаков, или, как говорят, упорядочить по Уолшу:

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}.$$

Каждой базисной функции Уолша, или базисному сигналу, сопоставляется строка в матрице Адамара, причём число перемен знаков — это аналог частоты этого сигнала.

Собственно преобразование заключается в том, что матрица Адамара размерности $2n$ умножается на столбец из $2n$ значений, соответствующих дискретизированному сигналу, в результате чего мы получаем вектор, компонентами которого являются коэффициенты разложения данного дискретного сигнала по дискретным функциям Уолша. В данном случае $2n$ — это количество временных отсчётов в соответствии с заданной дискретизацией.

Как определить частоту волны? Приближённая формула получается следующим образом. Берём коэффициент с наибольшим модулем и составляем пропорцию из условия, что 65536-ый коэффициент соответствует частоте 22050Hz:

$$\upsilon = \frac{k \cdot \upsilon_{\max}}{k_{\max}}.$$

Здесь υ — искомая частота, k — номер преобладающего коэффициента, $\upsilon_{\max}$ — максимальная частота, которую можно измерить (в зависимости от дискретизации), $k_{\max}$ — номер последнего коэффициента в буфере.

3.1. Анализ результатов (WAVE)

Как уже говорилось ранее, после того как преобразовали файлы из формата MIDI в формат WAVE, мы уже перестали иметь дело с символьными последовательностями. Теперь мы проанализируем те же мелодии, но с точки зрения спектрального анализа звуковых сигналов.

Сравнивая эталон и 1-ю мелодию, полученную из него вертикальным переносом на пять ступеней вверх, мы подвергаем преобразованию Адамара оба WAVE-файла. На рис. 1 изображена исходная мелодия после преобразования Адамара, стрелкой указан пик — преобладающий коэффициент (его значение — 5842).

Считаем по указанной выше формуле искомую частоту υ:

$$\upsilon = \frac{k \cdot \upsilon_{\max}}{k_{\max}},\ k = 5842,\ \upsilon_{\max} = 22050,\ k_{\max} = 262144.$$

Таким образом, искомая частота $\upsilon \approx 491.3$ Hz, что приблизительно соответствует частоте звука для ноты «си».

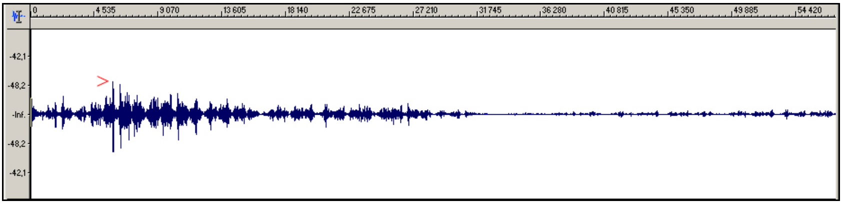

Рис. 1. Эталон после преобразования Адамара

Проделываем то же самое для 1-й трансформированной мелодии (рис. 2).

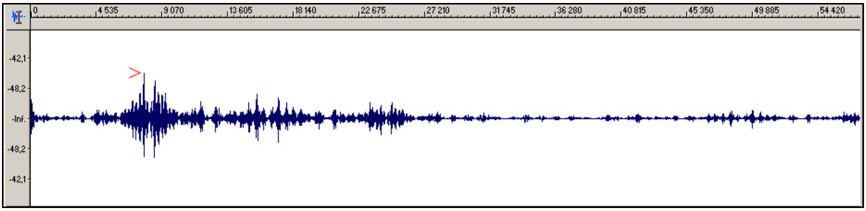

Рис. 2. Мелодия, трансформированная вертикальным переносом после преобразования Адамара

Теперь преобладающий коэффициент $k = 7829$, следовательно $\upsilon \approx 659$ Hz, что приблизительно соответствует частоте звука для ноты «ми».

Расстояние от ноты «си» вверх до ноты «ми» — как раз пять ступеней.

Посмотрим, как выглядит после преобразования Адамара 4-я трансформированная мелодия. Из всех четырех она максимально различается с эталоном.

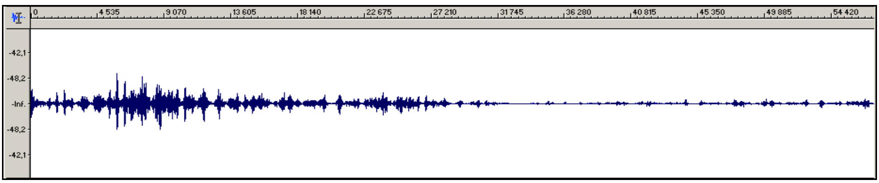

Рис. 3. 4-я трансформированная мелодия после преобразования Адамара

Здесь (на рис. 3) преобладающий коэффициент такой же, как и у эталона, поэтому общего переноса мелодии по вертикали вверх или вниз нет. Некоторые пики исчезли, некоторые добавились — это свидетельствует об исчезновении и появлении некоторых частот, более конкретные выводы можно будет сделать только при доскональном анализе.

4. ЗАКЛЮЧЕНИЕ

Спектральный анализ звуковых сигналов с помощью преобразования Адамара позволяет легко выявлять изменения тембра или пока только трансформации 1-го типа, более сложные трансформации анализу подвергаются с очень большим трудом. И это в случае одноголосых мелодий. Для монофонических мелодий задача определения тембра решаема, но, когда мы будем иметь дело с полифоническими композициями с несколькими тембрами, такой анализ произвести будет практически невозможно. Также спектральный анализ бессилен в тех случаях, когда трансформироваться будет только ритмическая часть.

Для определения структуры мелодии лучше всего подходить к анализу с точки зрения символьных последовательностей. Для получения более достоверных данных необходимо модифицировать метрику таким образом, чтобы учитывать музыкальные правила, т.е. разным преобразованиям ставить в соответствие разные цены.

СПИСОК ЛИТЕРАТУРЫ

1. **Зарипов Р. Х.** Машинный поиск вариантов при моделировании творческого процесса. — М.: Наука, 1983.
2. **Духнич Е. И.** Об одном подходе к выполнению цифровых линейных преобразований. — Кибернетика. — 1981. — № 5.
3. **Рабинер Л., Шафер Р.** Цифровая обработка речевых сигналов. — М.: Радио и связь, 1981.
4. **Algolist** http://algolist.manual.ru
5. **ITMAN** http://itman.narod.ru