

ИСПОЛЬЗОВАНИЕ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА ДЛЯ РАСПОЗНАВАНИЯ ЖАНРОВ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ

© 2016 О.П. Солдатова

Самарский национальный исследовательский университет имени академика С.П. Королёва

Статья поступила в редакцию 11.11.2016

В статье приводятся результаты решения задачи распознавания жанров музыкальных композиций при помощи модели многослойного перцептрона. Описана методика извлечения классификационных признаков из аудиофайлов музыкальных композиций, приведён перечень используемых в данной работе характеристик. Приведена структура многослойного перцептрона, описан алгоритм обучения. Исследована зависимость погрешности обучения многослойного перцептрона от значения коэффициента обучения и от числа нейронов в скрытом слое. Исследована точность распознавания музыкальных композиций, принадлежащих пяти различным жанрам на наборе музыкальных композиций GTZAN, состоящем из 1000 примеров.

Ключевые слова: классификационные признаки музыкальных композиций, распознавание жанра, быстрое преобразование Фурье, многослойный перцептрон, алгоритм наискорейшего спуска, метод обратного распространения ошибки.

ВВЕДЕНИЕ

В настоящее время становится весьма актуальной задача распознавания и поиска музыкальных композиций в связи с увеличением числа интернет сервисов с музыкальной тематикой и ростом их популярности. Поиск композиций можно разделить на два этапа. К первому этапу относится распознавание жанра музыкальной композиции с использованием модели многослойного перцептрона. Вторым этапом является сам поиск дубликата в рамках базы данных композиций найденного жанра.

Общую схему решения задачи распознавания жанров музыкальных произведений можно условно разделить на три этапа:

- определение набора классификационных признаков произведений;
- выделение признаков из файлов аудио-записей;
- классификация при помощи выбранной модели.

Задача точной идентификации жанра является сложной как для человека, так и для компьютера. Часто не существует общепринятого понятия классификационных признаков того или иного жанра и определения типов жанров. Несмотря на то, что разделение музыки по жанрам является в немалой степени субъективным, для описания того или иного жанра можно использовать критерии, связанные с фактурой, инструментовкой и ритмической структурой музыки. Наличие надёжных тестовых данных большого объёма также является ключевым требованием для эффективного обучения классификаторов музыкальных

произведений по жанрам, однако в открытом доступе практически нет подобных баз данных.

Нейросетевые классификаторы в настоящее время являются широко используемыми, ввиду их способности обучаться на ограниченном множестве примеров и наличии эффективных алгоритмов обучения. Одним из самых известных и хорошо изученных типов нейронных сетей является многослойный перцептрон. Обычно многослойная сеть состоит из множества входных узлов, которые образуют входной слой; одного или нескольких скрытых слоев вычислительных нейронов и одного выходного слоя. Входной сигнал распространяется по сети в прямом направлении от слоя к слою. Многослойные перцептроны обучаются с учителем, при этом обучение с учителем обычно выполняется с помощью градиентных алгоритмов и метода обратного распространения ошибки [1,2]. В данной работе приведены результаты решения задачи распознавания жанров музыкальных произведений с использованием модели многослойного перцептрона с использованием свободно распространяемой базы музыкальных композиций GTZAN.

1. КЛАССИФИКАЦИОННЫЕ ПРИЗНАКИ МУЗЫКАЛЬНЫХ ЖАНРОВ

Основной задачей анализа аудио сигналов является извлечение признаков, характеризующих сигнал. В большинстве работ предлагается три набора признаков, служащих для представления тембра, ритма и высоты звука [3]. Набор тембральных признаков совпадает с признаками распознавания речи и звуков в целом, однако два других набора характеризуют аспекты музыки. В литературе тембр определяется как характеристика звука, позволяющая двум звукам с одина-

Солдатова Ольга Петровна, кандидат технических наук, доцент, доцент кафедры информационных систем и технологий. E-mail: op-soldatova@yandex.ru

ковой высотой и громкостью звучать по-разному [3]. Для извлечения признаков, описывающих тембр, используется спектральное распределение сигнала, однако некоторые из таких признаков вычисляются во временной области. В литературе приводится список признаков, используемых для описания тембра инструментов [4]. Большинство из этих признаков можно использовать и в контексте распознавания музыкального жанра. Также в работе [4] приводятся основные низкоуровневые признаки, используемые в приложениях для определения жанров:

- временные признаки – вычисляются из кадров звукового сигнала;
- энергетические признаки – среднеквадратичная энергия кадра сигнала, энергия гармонической составляющей спектра сигнала, энергия шумовой части спектра и т. д.;
- спектральные признаки – признаки, описывающие форму спектра звукового кадра;
- перцептивные признаки – признаки, относящиеся к восприятию, вычисляются с использованием модели человеческого восприятия звука (относительная громкость, чёткость и т. д.).

Преобразования значений признаков, такие как производные первого или второго порядков, являются распространенным методом создания новых признаков или повышения размерности векторов признаков. В контексте задачи классификации значения тембральных признаков часто обобщаются с помощью применения статистик низкого порядка к более крупным окнам, называемым текстурными окнами [5]. Под признаками, описывающими высоту сигнала, подразумеваются признаки, описывающие гармонию и мелодию аудиозаписи музыкального произведения. Гармония может быть определена как использование и изучение слаженности звуков и аккордов. Мелодия же, напротив, представляет собой последовательность звуков различной высоты, воспринимаемых как единое целое. Гармонию иногда называют «горизонтальным» измерением музыки, а мелодию – «вертикальным» измерением. Гармония и мелодия более надёжно описываются низкоуровневыми атрибутами, нежели нотами или аккордами [5].

Главной идеей большинства анализаторов мелодии или гармонии является использование функции, описывающей распределение высот на коротком участке композиции. Высокоуровневые признаки тональность, основная частота, последовательность аккордов не используются, вместо этого на основе функции вычисляется набор значений амплитуд и расположения основных пиков, величин интервалов между пиками и любые другие статистические описатели распределения функции высотного содержимого.

Точного определения ритма не существует. Большинство авторов рассматривают ритм как идею временной регулярности. В более общем смысле, слово «ритм» может использоваться для

обозначения всех временных аспектов музыкального произведения. Ритмический рисунок является важной отличительной особенностью того или иного жанра [3]. Современные системы определения ритма имеют ряд недостатков, поэтому в системах автоматической классификации по жанрам чаще используется низкоуровневый подход [6]. Используя тот же подход, что и для низкоуровневых высотных признаков, можно извлечь признаки с помощью функции, оценивающей периодичности в диапазоне воспринимаемых темпов (обычно от 40 до 200 ударов в минуту). Признаки могут извлекаться из всего аудио сигнала, однако во многих задачах классификации используется небольшой музыкальный сегмент, в котором содержится достаточное количество информации для описания всего произведения, так как во многих музыкальных жанрах наблюдаются повторы музыкальной структуры. Кроме того, при использовании небольшой части сигнала можно значительно уменьшить вычислительные затраты. Довольно часто используется один 30-секундный отрезок, взятый через 30 секунд после начала композиции [3].

Тембр содержит достаточно информации для грубого определения музыкальных жанров; так, участники одного научного исследования, были способны выполнять правильную классификацию музыки по 10 жанрам в 53% случаев после прослушивания всего лишь 250 миллисекунд, и в 72% случаев после прослушивания 3 секунд записи. Это позволяет предположить, что для определения жанров не требуется высокоуровневого понимания музыки [7]. С помощью современного алгоритма определения сходства тембра было проведено исследование на базе данных из 20000 произведений в 18 жанрах [3]. Результаты показали, что между тембром и жанрами существует лишь небольшая зависимость, вследствие чего можно предположить, что схемы классификации, основанные только на тембре, ограничены по своей природе.

2. ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ ИЗ АУДИО ЗАПИСЕЙ

Методы для извлечения спектральных признаков, работают с амплитудным спектром сигнала. Спектр представляет собой массив комплексных чисел и вычисляется следующим образом:

- модуль обработки аудио сигнала считывает данные аудио файла, на вход подается имя аудиофайла, а на выходе модуля необработанные звуковые данные – ряд отсчётов, взятых через равные промежутки времени. Отсчёты представляют собой вещественные числа. Стандартная частота дискретизации для wave-файлов составляет 44100 Гц, соответственно, амплитуда сигнала измеряется 44100 раз в течение одной секунды. Звуковые каналы (если звук был многоканальным) при этом сводятся в один;

- при помощи быстрого преобразования Фурье все отсчеты представляются комплексным числом. Для массива комплексных отсчетов вычисляется амплитудный спектр звукового файла. Весь спектр разбивается на отдельные окна – кадры. Значения признаков, полученные для спектров отдельных кадров, нормализуются и усредняются. Затем из спектра извлекаются характеристики необходимые для классификации и поиска дубликата.

- Перечислим используемые в качестве классификационных признаков в данной работе характеристики сигнала:

- ZeroCrossingRate – характеристика отражающая частоту переходов аудио сигнала через ноль, то есть число изменений знака между идущими подряд значениями сигнала, делённое на общее число значений. Частота переходов через ноль является мерой шумности сигналов. Чтобы избежать малых колебаний значения величины около нуля, вызванных шумом, задаётся пороговое значение;

- DistributionShape – характеристика, показывающая значения дисперсии, коэффициентов асимметрии и эксцесса (остроты пика распределения) из массива центральных моментов распределения. Массив центральных моментов вычисляется с помощью алгоритма, который для заданного массива значений (в данном случае – спектра аудио сигнала) вычисляет центральные моменты от 0 до 4 порядка включительно;

- Root Mean Square – характеристика средне-квадратичного значения входного массива (в данном случае – спектра аудио сигнала);

- Spectral rolloff – характеристика спектрального спада для заданного спектра, которая определяется как частота, ниже которой сосредоточен заданный процент общей энергии спектра. По умолчанию используется значение 85%. Частота спектрального спада может использоваться для отделения шума от значимого содержимого;

- Strong-peak – характеристика «сильного пика» для заданного спектра, которая определяется как отношение между значением максимального пика спектра и «шириной» этого пика над порогом, равным половине амплитуды максимального пика. Это отношение показывает, есть ли в спектре «ярко выраженный» максимальный пик;

- Pitch-saliense – характеристика рельефности высоты для заданного спектра, которое определяется как отношение наибольшего значения автокорреляции спектра к несдвинутому значению автокорреляции. Немзыкальные звуковые эффекты и чистые тона (звуки, совершающие гармонические колебания одинаковой частоты) имеют значение рельефности высоты, близкое к нулю, в то время как звуки, спектр которых содержит несколько гармоник, имеют большее значение рельефности высоты;

- Volume – характеристика громкости аудио сигнала, которая определяется по закону Стивенса как энергия сигнала в степени 0.67. Для вычисления данного признака используются интервалы из 88200 отсчётов аудио сигнала (2 секунды при частоте дискретизации 44100 Гц) с размером скачка 44100 отсчётов. Значение громкости считается на каждом таком интервале, после чего вычисляется среднее значение громкости;

- Beats-per-minute – характеристика темпа аудио сигнала, выраженная в ударах в минуту.

Таким образом, каждая аудиозапись описывается 8 характеристиками, которые подаются на вход нейронной сети.

3. КЛАССИФИКАЦИЯ МУЗЫКАЛЬНЫХ КОМПОЗИЦИЙ ПО ЖАНРАМ ПРИ ПОМОЩИ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА

С точки зрения современного программного обеспечения, ни одна система распознавания музыкальных композиций не достигла достаточно высоких показателей для реального применения. Например, наиболее высокий процент успешной классификации на MIREX 2013 (ежегодное соревнование по различным областям музыкального информационного поиска, в том числе классификации произведений по жанрам) составил 76% при классификации по 10 жанрам [8].

Компетентность и затраты времени, требуемые для ручной классификации записей, создают серьёзные препятствия для получения качественных тестовых данных. Это особенно заметно, когда для обучения требуются большие наборы данных. Ещё больше усложняет ситуацию появление новых жанров и изменение понимания уже существующих, что может потребовать повторного обучения и новой аннотации контрольных данных. Необходимость большой обучающей выборки также имеет значение с точки зрения машинного обучения. Мощные алгоритмы обучения, такие как метод опорных векторов или AdaBoost эффективно моделируют сложные пространства жанров, но многие из таких алгоритмов плохо масштабируются. Так же алгоритмы бустинга могут приводить к построению громоздких композиций, которые исключают возможность содержательной интерпретации, требуют больших объёмов памяти для хранения базовых алгоритмов и существенных затрат времени на вычисление классификаций [9]. Вследствие приведённых выше причин, в работе [10] для классификации музыкальных композиций была использована модель одно-слойного перцептрона, а в данной работе для классификации жанров была использована модель многослойного перцептрона.

Архитектура двухслойной сети представлена на рис. 1.

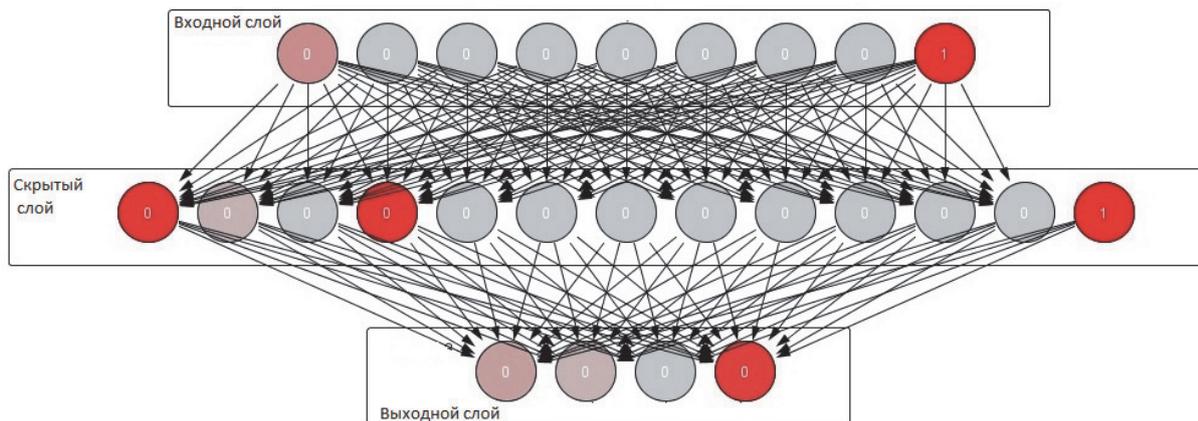


Рис. 1. Архитектура многослойного перцептрона

Входной слой не является слоем нейронов, а представляет собой входной вектор характеристик музыкальных композиций, представленных действительными числами. Нейроны скрытого и выходного слоёв используют сигмоидальные функции активации.

На сегодняшний день многослойный перцептрон - одна из самых популярных и используемых моделей нейронных сетей [1]. Одно из главных преимуществ многослойного перцептрона - это возможность решать алгоритмически неразрешимые задачи или задачи, для которых алгоритмическое решение неизвестно, но для которых возможно составить репрезентативный набор примеров с известными решениями. При обучении многослойный перцептрон за счёт своего внутреннего строения, выявляет закономерности в связи входных и выходных образов, тем самым как бы «обобщает» полученный на обучающей выборке опыт. В этой способности к обобщению и состоит основа привлекательности многослойного перцептрона. В данной работе для обучения сети используется алгоритм наискорейшего спуска и метод обратного распространения ошибки.

Входной слой многослойного перцептрона представляет собой входной вектор в виде $x = [x_0, x_1, \dots, x_N]^T$. С вектором x связаны два выходных вектора сети: вектор фактических выходных сигналов $y = [y_0, y_1, \dots, y_M]^T$ и вектор ожидаемых выходных сигналов $d = [d_0, d_1, \dots, d_M]^T$.

Обозначим вектор весов скрытого слоя как $w_{ij}^{(1)}$, а вектор весов выходного слоя как $w_{si}^{(2)}$, где $j = 0, 1, \dots, N; i = 0, 1, \dots, K; s = 1, 2, \dots, M$, N - размер входного вектора, K - число нейронов в скрытом слое, M - число нейронов в выходном слое. Если обозначить функцию активации нейронов как f , то s -ый нейрон выходного слоя вырабатывает выходной сигнал, определяемый следующей формулой:

$$y_s = f\left(\sum_{i=0}^K w_{si}^{(2)} v_i\right) = f\left(\sum_{i=0}^K w_{si}^{(2)} f\left(\sum_{j=0}^N w_{ij}^{(1)} x_j\right)\right). \quad (1)$$

Цель обучения состоит в подборе таких значений весов, чтобы при заданном входном векторе x получить на выходе значения сигналов y_s , которые с требуемой точностью будут совпадать с ожидаемыми значениями d_s .

При обучении ставится задача минимизации целевой функции, которая для P обучающих выборок определяется по формуле:

$$E(w) = \frac{1}{2} \sum_{t=1}^P \sum_{s=1}^M \left(y_s^{(t)} - d_s^{(t)}\right)^2. \quad (2)$$

Уточнение весов может проводиться после предъявления каждой обучающей выборки.

В данной работе для обучения многослойного перцептрона был использован алгоритм наискорейшего спуска и метод обратного распространения ошибки.

В алгоритме наискорейшего спуска формулу уточнения весов можно представить следующим образом:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \cdot \frac{\partial E(t)}{\partial w_{ij}(t)}, \quad (3)$$

где η - коэффициент обучения, $0 < \eta < 1$.

Обучение многослойной сети с применением градиентных методов требует определения вектора градиента относительно весов всех слоев сети. Эта задача имеет очевидное решение только для весов выходного слоя. Для других слоев используется метод обратного распространения ошибки. Обучение сети с использованием метода обратного распространения ошибки проводится в несколько этапов.

На первом при предъявлении входного вектора x рассчитываются вначале значения выходных сигналов v_i нейронов скрытого слоя, а затем значения y_s нейронов выходного слоя. После получения значений выходных сигналов y_s становится возможным рассчитать фактическое значение целевой функции ошибки $E(w)$. На втором этапе минимизируется значение этой функции.

Рассмотрим основные расчетные формулы для сети с одним скрытым слоем, представленной на рис. 1. С учетом принятых обозначений

целевая функция для выходного слоя нейронов определяется следующим образом:

$$E = \frac{1}{2} \sum_{s=1}^M \left[f \left(\sum_{i=0}^K w_{si}^{(2)} v_i \right) - d_s \right]^2 = \frac{1}{2} \sum_{s=1}^M \left[f \left(\sum_{i=0}^K w_{si}^{(2)} f \left(\sum_{j=0}^N w_{ij}^{(1)} x_j \right) \right) - d_s \right]^2, \quad (4)$$

$$\frac{\partial E}{\partial w_{si}^{(2)}} = (y_s - d_s) \cdot \frac{df(u_s^{(2)})}{du_s^{(2)}} \cdot v_i. \quad (5)$$

В формуле (5) и под переменной u будем понимать выходные сигналы сумматоров нейронов скрытого или выходного слоя, представленных формулами (6) и (7).

$$u_i^{(1)} = \sum_{j=0}^N w_j^{(1)} x_j, \quad (6)$$

$$u_s^{(2)} = \sum_{i=0}^K w_{si}^{(2)} v_i. \quad (7)$$

Компоненты градиента относительно нейронов скрытого слоя описываются более сложной зависимостью:

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = \sum_{s=1}^M (y_s - d_s) \cdot \frac{dy_s}{dv_i} \cdot \frac{dv_i}{dw_{ij}^{(1)}}. \quad (8)$$

В другом виде эта зависимость может быть выражена формулой:

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = \sum_{s=1}^M (y_s - d_s) \cdot \frac{df(u_s^{(2)})}{du_s^{(2)}} \cdot w_{si}^{(2)} \cdot \frac{df(u_i^{(1)})}{du_i^{(1)}} x_j. \quad (9)$$

4. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

В качестве проверочных данных используется набор музыкальных композиций GTZAN, состоя-

Таблица 1. Зависимость СКО погрешности обучения от значения коэффициента обучения

Значение коэффициента обучения	СКО погрешности обучения				
	Блюзовые композиции	Классические композиции	Хип-хоп композиции	Джазовые композиции	Рок композиции
0,10	0,0328	0,0189	0,0874	0,0430	0,0231
0,20	0,0128	0,0124	0,2290	0,0491	0,0511
0,30	0,0031	0,0017	0,0920	0,0376	0,0124
0,40	0,0030	0,0018	0,0140	0,0094	0,0052
0,41	0,0024	0,0015	0,0098	0,0092	0,0049
0,42	0,0019	0,0012	0,0090	0,0091	0,0047
0,43	0,0062	0,0031	0,0106	0,0095	0,0108
0,45	0,0137	0,0073	0,0129	0,0099	0,0194
0,50	0,0243	0,0108	0,0143	0,0108	0,0227
0,60	0,0275	0,0053	0,0090	0,0664	0,0490
0,70	0,0276	0,0107	0,0157	0,0125	0,0072
0,80	0,0038	0,0073	0,0091	0,0064	0,0125
0,90	0,0312	0,0213	0,0774	0,0230	0,0230

щий из 1000 примеров. Для проверки работы этот набор был разбит на обучающую выборку, состоящую из 900 примеров (по два набора по 90 примеров на каждый из пяти жанров) и проверочную выборку, состоящую из 100 примеров (по два набора по 10 примеров на каждый жанр). Использование 10% данных в качестве тестовых данных является распространённой практикой. Целью проведённых исследований является определение качества обучения нейронной сети и качества классификации, а также выявление влияния коэффициента обучения и параметров сети на погрешность обучения. Размер входного слоя сети определяется количеством характеристик музыкальных композиций, в данной работе использовались 8 признаков. Размер выходного слоя определяется числом распознаваемых жанров, в данном исследовании распознавались музыкальные произведения, принадлежащие к 5 различным жанрам. Поэтому при проведении исследований в качестве параметра сети, влияющего на погрешность обучения, было взято число нейронов скрытого слоя.

Как показатель качества обучения для количественной оценки этой разности будем рассчитывать, в дополнение к значению функции ошибки, определяемому формулой (2), суммарное среднеквадратическое отклонение (СКО) погрешности обучения:

$$CKO = \sqrt{\frac{1}{P-1} \cdot \frac{1}{M} \cdot \sum_{t=1}^P \sum_{s=1}^M (y_s^{(t)} - d_s^{(t)})^2}. \quad (10)$$

На этапе тестирования сети будем считать процент верно распознанных векторов среди общего числа тестирующих данных. В дальней-

шем будем называть эту оценку качества точностью классификации. Точность классификации характеризует эффективность решения задачи классификации. В табл. 1 приведены результаты исследования зависимости СКО погрешности обучения от значения коэффициента обучения.

Как видно из таблицы наименьшие значения СКО наблюдаются при значении коэффициента обучения равном 0,42. В таблице 2 приведены результаты исследования зависимости СКО погрешности обучения от числа нейронов в скрытом слое.

Как видно из таблицы наименьшие значения СКО наблюдаются при числе нейронов равном 23. В табл. 3 приведены результаты исследования точности классификации музыкальных жанров на тестовых наборах.

ЗАКЛЮЧЕНИЕ

Из табл. 3 видно, что средняя точность классификации композиций, равная 71% является приемлемой. В работе 10 приведены данные о точности автоматической классификации порядка 80%, однако при этом использовалось значительно большее число характеристик музыкальных композиций (порядка 90). Таким образом, можно сказать, что многослойный персептрон достаточно эффективно решает задачу классификации музыкальных композиций по нескольким характеристикам.

СПИСОК ЛИТЕРАТУРЫ

1. Осовский С. Нейронные сети для обработки информации [пер. с польского И.Д. Рудинского]. М.:

Таблица 2. Зависимость СКО погрешности обучения от числа нейронов в скрытом слое

Количество нейронов в скрытом слое	СКО погрешности обучения				
	Блюзовые композиции	Классические композиции	Хип-хоп композиции	Джазовые композиции	Рок композиции
20	0,0638	0,0772	0,0972	0,0555	0,0419
21	0,0574	0,0224	0,2290	0,0491	0,0511
22	0,0277	0,0676	0,0860	0,0376	0,0124
23	0,0120	0,0216	0,0460	0,0125	0,0152
24	0,0146	0,0218	0,0523	0,0180	0,0127
25	0,0132	0,0224	0,0520	0,0180	0,0186

Таблица 3. Точность классификации

Входные данные	Точность классификации музыкальных жанров				
	Блюзовые композиции	Классические композиции	Хип-хоп композиции	Джазовые композиции	Рок композиции
Набор 1 рок композиций	7%	10%	5%	7%	75%
Набор 2 рок композиций	10%	8%	5%	6%	71%
Набор 1 джазовых композиций	10%	10%	0%	77%	3%
Набор 2 джазовых композиций	10%	10%	5%	70%	5%
Набор 1 хип-хоп композиций	3%	8%	75%	4%	10%
Набор 2 хип-хоп композиций	7%	0%	78%	5%	10%
Набор 1 классических композиций	8%	69%	6%	10%	7%
Набор 1 классических композиций	11%	67%	4%	10%	8%
Набор 1 блюзовых композиций	60%	3%	0%	30%	7%
Набор 2 блюзовых композиций	68%	4%	3%	15%	10%

- Финансы и статистика, 2002. 344 с.
- Haykin S.S. et al. Neural networks and learning machines // Upper Saddle River: Pearson Education, 2009. V. 3.
 - Scaringella N., Zoia G., Mlynek D. Automatic genre classification of music content: a survey // Signal Processing Magazine, IEEE (Volume 23, Issue 2), 2006. Pp. 21.
 - Peeters G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project . CUIDADO I.S.T. Project Report, 2004. Pp. 25.
 - Meng A. Ahrendt P., Larsen J. Improving Music Genre Classification by Short-Time Feature Integration // IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Pp 4.
 - Gouyon F. Dixon S. A review of automatic rhythm description system // Computer Music Journal. 2005. Vol. 29. Pp. 34-54.
 - Aucouturier J.J., Pachet F. Representing musical genre: a state of the art // Journal of New Music Research. 2003. Vol. 32. No. 1. Pp. 83-93.
 - MIREX 2013: Audio Train Test: Genre Classification (Mixed) - MIREX08 Dataset [Электронный ресурс]/ URL: http://www.music-ir.org/nema_out/mirex2013/results/act/mixed_report/summary.html (дата обращения 07.09.2016).
 - MachineLearning.Ru — информационно-аналитический ресурс по машинному обучению [Электронный ресурс]. URL: <http://www.machinelearning.ru/> (дата обращения 07.09.2016).
 - Солдатова О.П., Мушин А.Е. Классификация музыкальных композиций с использованием однослойного перцептрона // Труды международной научно-технической конференции «Перспективные информационные технологии (ПИТ 2016)» (Самара, СГАУ, 26–28 апреля 2016 г.). Самара: Издательство Самарского научного центра РАН, 2016. С.355-358.

USING THE MULTI-LAYER PERCEPTRON FOR RECOGNIZING THE MUSIC GENRES

© 2016 O.P. Soldatova

Samara National Research University named after Academician S.P. Korolyov

The article shows the results of recognizing the music genres by using the multi-layer perceptron. It describes a method to extract classifying parameters from audio files with music compositions, shows a list of characteristics used in the article. Multi-layer perceptron structure and its learning algorithm are both described. The learning error shows a dependency from a value of the learning coefficient and a number of hidden neurons. Recognizing accuracy is tested on a set of music compositions related to five different genres GTZAN containing 1000 examples. *Keywords:* classifying parameters of music compositions, genres recognition, fast Fourier transform, multi-layer perceptron, fast descending algorithm, backpropagation algorithm.