

Е.Ю. КОСАРЕВ, студент НТУ «ХПИ»

ИДЕНТИФИКАЦИЯ ЗВУКОВЫХ СИГНАЛОВ

Статья посвящена изложению способа распознавания звуковых сигналов нейронной сетью с использованием спектрального анализа методом Монте-Карло, что позволяет уменьшить вычислительные затраты по сравнению с классическим спектральным анализом.

Стаття присвячена викладенню способу розпізнавання звукових сигналів нейронною мережею з використанням спектрального аналізу методом Монте-Карло, що надає можливість зменшити обчислювальні витрати у порівнянні з класичним спектральним аналізом.

Article is dedicated to method of identification of sound signals by neural network using spectral analysis with Monte-Carlo. This provides decreasing of calculations as compared with classical spectral analysis.

Создание высокоавтоматизированных трудосберегающих производств предполагает автоматизацию не только физического, но и интеллектуального труда человека. Автоматизация интеллектуальной деятельности потребовала решения ряда новых задач, не возникавших ранее. К их числу относятся описание и представление в ЭВМ сложной внешней среды, автоматическое планирование и выполнение комплекса разнообразных действий технического устройства, направленных на достижение заданной цели, организация общения человека с ЭВМ на языке, близком к естественному, и ряд других.

Идентификация звуковых сигналов относится к разряду интеллектуальных инженерных задач и является одной из наиболее перспективных. Область применения системы, решающей данную задачу, довольно широка – это робототехника, системы управления ЭВМ с помощью речи, криминалистика, системы безопасности и др.

В данной статье описан математический аппарат системы распознавания звуковых сигналов, которая программно реализуется на ЭВМ. Рассматриваются звуковые сигналы, продолжительность звучания которых ограничена 6-10 секундами и имеется обучающая выборка – шаблон, в котором образцам звуковых сигналов соответствуют классы, к которым они принадлежат. Требуется идентифицировать неизвестный звуковой сигнал – отнести его к одному из классов, существующих в шаблоне, либо отнести к классу неизвестных.

Процесс физиологического распознавания речи может быть промоделирован с помощью электронных устройств (рис. 1), где каждому этапу обработки информации в живом организме соответствует определенная техническая операция [1].



Рисунок 1 – Этапы распознавания звуковых сигналов

Выбор типа датчика зависит от ряда условий: наличие шума, требования к направленности и т. д. Спектральный анализ позволяет определить функцию, выполняемую ушной улиткой, сводящуюся к определению интенсивности сигнала в зависимости от частоты. Выделение свойств, т.е. характерных элементов, осуществляют совместно улитка, слуховой нерв и кортиев орган. Наконец, на верхнем уровне, в коре головного мозга, происходит понимание сообщения.

Как видно, реализация системы распознавания звуковых сигналов сводится к следующим подзадачам: разработка и реализация процедуры спектрального анализа, разработка и выбор метода выделения свойств на основе спектра, реализация процедуры распознавания.

Сигнал, записанный через микрофон, подключенный к звуковой карте из аналогового преобразуется в цифровой путем дискретизации. Это реализовано на низком уровне – аппаратном и системном уровне драйверов звуковой карты. Вследствие дискретизации вектор, описывающий сигнал имеет очень большую размерность. Так при частоте дискретизации 11025 Гц десять секунд звукового сигнала будет описывать вектор из 110250 значений, при чем надо учитывать, что частота дискретизации 11025 Гц довольно невысокий показатель для современной цифровой звукозаписи. Это значит, что вектор, описывающий такой небольшой промежуток времени, как десять секунд, может иметь и большую размерность. Для систем распознавания векторы таких размерностей неприемлемы, так как распознавание и обучение будут требовать очень много вычислительного времени и ресурсов. Следовательно, исходная задача разбивается на две подзадачи – выделение «квинтэссенции» сигнала и её последующее распознавание. Под «квинтэссенцией» будем понимать некий набор свойств (признаков),

объединяющий сигналы в общее множество – образ. Каждое отдельное проявление из множества объединенных в образ будем называть изображением.

Как было сказано выше, дискретизированный звуковой сигнал требует некоторого преобразования, которое выделит его свойства, на основании которых, сигналы можно будет объединять в образы. Для субъективного распознавания любого звука основную роль играют две его характеристики – высота и тембр [1].

Высота характеризует основную частоту звука.

Тембр характеризует индивидуальную «окраску» звука, присущие только ему особенности, позволяющие отличить данный конкретный звук от других, имеющих такую же интенсивность и частоту. Тембр определяется наличием гармоник, чья амплитуда должна быть достаточно заметной, чтобы внести «вклад» в восприятие звука [1]. Для выделения этих характеристик необходимо произвести спектральный анализ сигнала – выделить его спектр. Спектральный анализ – это один из методов обработки сигналов, который позволяет охарактеризовать частотный состав измеряемого сигнала. Преобразование Фурье является математической основой, которая связывает временной или пространственный сигнал (или же некоторую модель этого сигнала) с его представлением в частотной области. Любую периодическую функцию $f(t)$ с периодом T , удовлетворяющей условиям Дирихле, можно представить с помощью тригонометрического ряда Фурье:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\omega_1 t + b_k \sin k\omega_1 t), \quad (1)$$

где ω_1 – частота повторения (или частота первой гармоники);

k - номер гармоники

Этот ряд содержит бесконечное число косинусных или синусных составляющих – гармоник, причем амплитуды этих составляющих a_k и b_k являются коэффициентами Фурье, определяемыми интегральными выражениями:

$$a_k = \frac{2}{T} \int_0^T f(t) \cos k\omega_1 t dt, \quad (2)$$

$$b_k = \frac{2}{T} \int_0^T f(t) \sin \omega_1 t dt. \quad (3)$$

Таким образом, выделение свойств сигнала сводится к выделению спектра сигнала с и его последующим огрублением – формированием «квинтэссенции» спектра.

Численный спектральный анализ заключается в нахождении коэффициентов $a_0, a_1, \dots, a_{N-1}, b_0, b_1, \dots, b_{N-1}$ (или $A_0, A_1, \dots, A_{N-1}, \varphi_0, \varphi_1, \dots, \varphi_{N-1}$) для периодической функции $f(t)$ заданной на отрезке $[0, T]$ N дискретными отсчетами. Он сводится к вычислению (2), (3) по формулам численного интегрирования для метода прямоугольников [2]:

$$a_k = \frac{2}{N} \sum_{n=0}^{N-1} f(n) \cos(2\pi \frac{1}{N} kn), \quad (4)$$

$$b_k = \frac{2}{N} \sum_{n=0}^{N-1} f(n) \sin(2\pi \frac{1}{N} kn), \quad (5)$$

где учтено, что $f_1 = \frac{1}{T}$, шаг дискретизации по оси x равен $\Delta x = \frac{T}{N}$ и

соответственно $2\pi f_1 n \Delta x = 2\pi \frac{1}{N} kn$.

Найденные по формулам (4), (6) коэффициенты Фурье приближают функцию $f(t)$ рядом (1) с меньшей квадратичной ошибкой [3]. Данный метод имеет очень большой недостаток – невысокая скорость из-за большого количества вычислений.

Получить спектр функции $f(t)$ можно путем вычисления тех же (2) и (3), но в качестве метода интегрирования использовать метод Монте-Карло (метод статистических испытаний) с количеством испытаний меньше, чем N . Количество вычислений будет меньше, а, следовательно, скорость нахождения спектра выше.

Выбирается мажоранта – прямоугольник, стороны которого параллельны осям, а координаты вершин – $(0, f_{abs\ max}(t))$, $(0, -f_{abs\ max}(t))$, $(T, f_{abs\ max}(t))$, $(T, -f_{abs\ max}(t))$, где $f_{abs\ max}(t)$ – максимальное по модулю значение функции $f(t)$ на промежутке $[0, T]$. Выбор такой мажоранты обоснован тем, что подынтегральные выражения содержат $f(t)$ умноженную на косинус или синус. Следовательно, подынтегральная функция не выходит за рамки $[-f_{abs\ max}(t), f_{abs\ max}(t)]$, т.к. косинус и синус не превосходят по модулю 1. Проводится K статистических испытаний, в ходе которых генерируется две случайных величины. Одна равномерно распределенная на промежутке $[0, T]$, другая на промежутке $[-f_{abs\ max}(t), f_{abs\ max}(t)]$. В случае, если сгенерированная случайная точка находится между значением функции и осью x счетчик «удачных» испытаний L , изначально равный 0, увеличивается или уменьшается на единицу, в зависимости от знака функции. Таким образом, значения интегралов (2), (3) вычисляются по формуле:

$$I = S_{mj} \frac{L}{K}, \quad (6)$$

$$S_{mj} = T(f_{\max}(t) - f_{\min}(t)). \quad (7)$$

Полученные значения по (6) умножаются на $\frac{2}{T}$ в соответствии с (2), (3).

Точность и скорость работы метода зависят от количества испытаний K . При чем скорость зависит прямопропорционально, а точность обратнопропорционально.

Для подсчета точности работы метода с использованием интегрирования методом Монте-Карло использовался функционал невязки:

$$\Omega = \frac{\sum_{i=0}^R |\bar{s}_i - s_i|}{R}, \quad (8)$$

где R – размерность массива, описывающего спектр (массива амплитуд) в нашем случае $R = 5492$;

\bar{s} – массив амплитуд, полученный методом Монте-Карло;

s – массив амплитуд, полученный классическим спектральным анализом.

Так как метод Монте-Карло статистический, то для построения графика использовались средние значения ошибки. Изменение точности работы метода на рис. 2

Из приведенного выше следует, что последовательный численный спектральный анализ с использованием интегрирования по методу прямоугольников хоть и дает высокую точность, но требует намного больше вычислений и времени, чем метод Монте-Карло. Метод Монте-Карло требует намного меньше времени, но при этом вносится погрешность. Время работы метода возрастает линейно от количества испытаний, а ошибка уменьшается нелинейно. Цель данного звена системы не точное вычисление спектра, а выделение характерных признаков сигнала. С учетом того, что для распознавания важна скорость работы метода, можно сделать вывод, что метод Монте-Карло является наиболее подходящим.

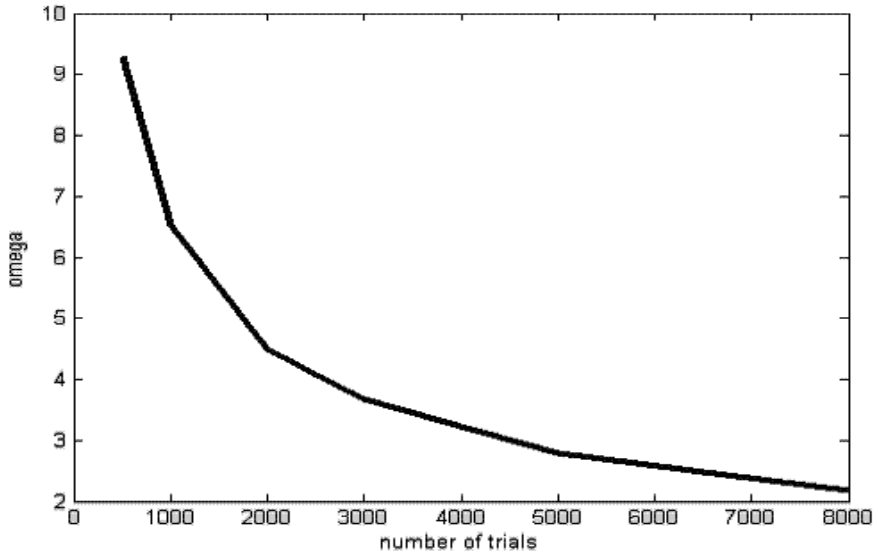


Рисунок 2 – Изменение ошибки вычисления спектра с использованием Монте-Карло от количества испытаний.

Рассмотрим методы выделения свойств из полученного спектра.

Вектор $s = (A_1, A_2, \dots, A_K)$, описывающий спектр, также имеет большую размерность. Выделить свойства сигнала на основе спектра можно несколькими способами: выделить M больших гармоник (рис. 3), составить вектор средней спектральной плотности по полосам частот (рис. 4) с количеством полос L и их шириной $\psi(l)$ (постоянной или переменной), где l – номер полосы.

Алгоритм метода выделения больших гармоник заключается в последовательном прохождении вектора S , описывающего спектр, с запоминанием частот и амплитуд больших гармоник в отдельном векторе.

Алгоритм вычисления вектора средней спектральной плотности по полосам частот заключается в последовательном прохождении вектора, описывающего спектр S , с накоплением сумм в каналах k_1, k_2, \dots, k_L и последующем делении значений k_1, k_2, \dots, k_L на соответствующие $\psi(l)$. Смена канала происходит при переходе в следующую полосу.

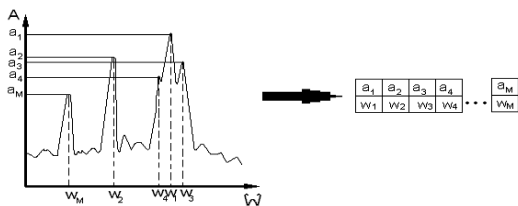


Рисунок 3 – Выделение M больших гармоник.

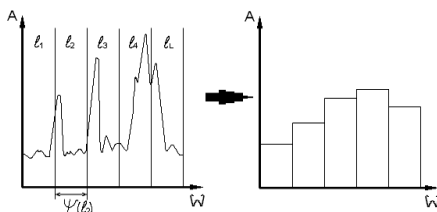


Рисунок 4 – Составление вектора средней спектральной плотности по полосам частот.

Для распознавания использовалась двухслойная нейронная сеть прямого распространения со входным слоем, содержащим l нейронов (по количеству признаков) и варьируемым выходным слоем (в зависимости от количества распознаваемых образов). Т.к. на выходе нейрон выдает ноль или один, то количество нейронов выходного слоя можно описать формулой:

$$N = \log_2 K, \quad (9)$$

где N – число нейронов;

K – количество распознаваемых образов.

Схема нейрона



Рисунок 5 – Структура искусственного нейрона.

Входной оператор преобразует взвешенные входы и подает их на оператор активации. В качестве входного оператора использовалась сумма взвешенных входов:

$$f(x, w) = \sum_{i=1}^N w_i x_i . \quad (13)$$

В качестве функции активации – функция единичного скачка.

Для нейронов первого слоя входными сигналами является взвешенная сумма вектора признаков. Входными сигналами нейронов второго слоя являются выходные сигналы предыдущего слоя (рис. 6).

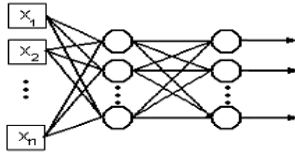


Рисунок 6 – Архитектура нейронной сети

Особенностью данной сети является способность к обучению ее весовых коэффициентов. В режиме обучения ей предъявляют образы – обучающие пары (x, y^*) , на основании которых она настраивает свои параметры так, чтобы при появлении некоторого входного вектора x на его выходе появлялись соответствующие этому входу сигналы y . Процесс обучения заканчивается, когда все пары (x, y^*) ассоциируются правильно. Алгоритм обучения нейронной сети реализует способ обучения с учителем, при котором ошибка минимизируется [4].

Компьютерная программа, реализующая данный метод, может распознавать сигналы от различных источников. Тестовым примером является распознавание звуков флейты, электрогитары и классической гитары. После обучения программа может распознавать неизвестные сигналы: относить их к одному из трех известных классов или к классу «неизвестный сигнал».

Данная работа является логически законченным исследованием, но может быть использована как база для последующих исследований в данном направлении.

Список литературы: 1.Фор А. Восприятие и распознавание образов – М: Машиностроение, 1989. 2.Мазманишвили А.С., Шкварко Ю.В. Методические указания к лабораторным занятиям по курсу «Численные методы», часть 2: «Численные методы прикладного анализа» – Х: ХПИ, 1992 3.Корн Г., Корн Т. Справочник по математике – М: Наука, 1973. 4. Руденко О.Г., Бодянский Е.В. Основы теории искусственных нейронных сетей – Х: ТЕЛЕТЕХ, 2002.

Поступила в редколлегию 27.04.07