

ющие значения: $\alpha = 0,8$; $\beta = 0,2$; количество используемых фильтров в банке равно 32. Фраза на русском языке: «штурман просил продолжить разворот».

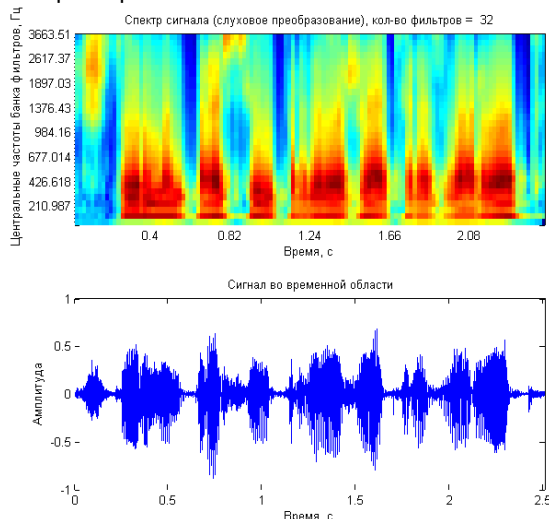


Рис.3 – Спектр сигнала, полученный с помощью слухового преобразования

Данное преобразование моделирует импульсные характеристики базилярной мембраны и её особенности нелинейного распределения частоты. Рассмотренное преобразование имеет следующие преимущества: устойчивость к шумовой составляющей сигнала, не имеет четко выраженных гармоник и также не даёт вычислительного шума. Эти преимущества могут быть использованы в различных приложениях, как в задачах идентификации диктора для получения характеристического вектора, так и для шумоподавления, синтеза музыки и речи и т.д. В области идентификации диктора, результатом слухового преобразования является сигнал, разложенный на частотные полосы, так как это делает БМ. Далее по аналогии со слуховой системой человека моделируется поведение волосковых клеток (внутренних и внешних) и учёт их нелинейности восприятия. После, следует этап уменьшения размерности полученных характеристик. Полученный характеристический вектор, называют кохлеарными кепстральными коэффициентами (ККК).

Таким образом, было рассмотрено и реализовано с помощью среды MATLAB слуховое преобразование. Были подобраны наиболее оптимальные параметры используемого банка кохлеарных фильтров для их дальнейшего использования в задачах распознавания. Полученные с помощью слухового преобразования характеристические признаки будут более устойчивы к изменению акустической среды в процессе распознавания, что повысит эффективность работы систем распознавания в акустических шумах.

Список использованных источников:

1. B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Amer., vol. 55, pp. 1304–1312, 1974.
2. Q. Li, An auditory-based transform for audio signal processing / Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust., – New Paltz, NY, 2009.

ОБУЧЕНИЕ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жук И.Н.

Азаров И.С. – к-т. техн. наук, доцент

В современных автоматизированных системах взаимодействия между человеком и вычислительной техникой все чаще встречается голосовая составляющая. Голосовое управление во многом позволяет убрать преграды непосредственного механического взаимодействия с устройствами. Реализация голосового командного интерфейса производится в большинстве случаев с помощью различных нейронных сетей.

Применение рекуррентной нейронной сети подразумевает использование выходных значений сети в качестве входных, что позволяет сохранять внутри сети информацию о предыдущих входных значениях. Данный подход позволяет представлять звук естественным последовательным образом. Подобная структура сети предполагает использование соответствующих обучающих алгоритмов. Для изучения был выбран алгоритм обратного распространения ошибки во времени. Суть алгоритма заключается в том, что мы разворачиваем рекуррентную сеть в очень длинную обычную сеть прямого распространения.

На рисунках 1 и 2 приведены структурные схемы формирования и обработки:

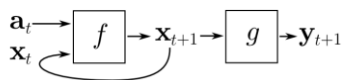


Рис. 1 – Структурная схема рекуррентной сети

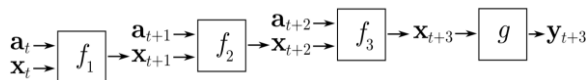


Рис. 2 – Структурная схема развернутой рекуррентной сети

Выходная ошибка рассчитывается по формуле: $E = \sum_j \frac{1}{2} (t_j - y_j)^2$,

Где E – ошибка, t – ожидаемое выходное значение, y – реальное выходное значение, суммирование выполняется по всем выходным нейронам сети.

$$\frac{\partial E}{\partial z_j} = \frac{\partial y_j}{\partial z_j} \frac{\partial E}{\partial y_j}$$

производная ошибки по суммарному входу каждого выходного нейрона

$$\frac{\partial E}{\partial y_i} = \sum_j w_{ij} \frac{\partial E}{\partial z_j}$$

производная ошибки по выходам нейронов предыдущего слоя

$$\frac{\partial E}{\partial w_{ij}} = y_i \frac{\partial E}{\partial z_j}$$

производная ошибки по весовым коэффициентам выходного нейрона

Где z_j – вход одного выходного нейрона, y_j – выход нейрона из предыдущего слоя, w_{ij} – весовой коэффициент нейрона выходного слоя, связывающий его с выходом из предыдущего.

Данная процедура выполняется через всю развернутую сеть от конца к началу. Полученные результаты приводят к следующим выводам: рекуррентные сети отлично подходят для обработки различного рода последовательностей (в данном случае звуковых). Однако их обучение весьма сложная нетривиальная задача, требующая глубокого изучения.

Для создания моделей была использована система компьютерной алгебры Wolfram Mathematica и Mathworks MATLAB. В качестве данных были использованы звуковые записи реальных акустических обстановок. Для обработки нейронной сетью из данных были извлечены параметрические векторы представляющие собой Мел-Кепстральные коэффициенты.

Основной сложностью при создании модели обработки стала обработка обратной связи, т.е. процесс разворачивания рекуррентной нейронной сети в сеть прямого распространения.

Таким образом было произведено обучение рекуррентной нейронной сети для распознавания речевых команд. Данная система за счет обратной связи адаптируется под различные временные нестационарности сигнала, а так же позволяет обрабатывать различную скорость произношения. Данная особенность выгодно выделяет подобный подход на фоне сетей прямого распространения.

Список использованных источников:

1. Лайонс, Р. Цифровая обработка сигналов / Р. Лайонс // Цифровая обработка сигналов: Второе издание. Пер. с англ. – М.: ООО «Бином-Пресс», 2006г. – 656с.:ил.
2. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82-97.

РАСПОЗНАТЕЛЬ РУКОПИСНЫХ ЦИФР НА ОСНОВЕ НЕЙРОННОЙ СЕТИ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Кустов А.Ю.

Вашкевич М.И. – канд. техн. наук, доцент

Задача по распознаванию рукописных цифр широко распространена и является активной темой в приложениях (ОПС) и в исследованиях по классификации/обучению образов. В ОПС приложениях распознавание цифр применяется для сортировки почтовых отправлений, обработки банковских чеков, ввода данных с форм и т.д. Для этих приложений производительность (точность и скорость) распознавания цифр имеет решающее значение для эффективной работы. Среди разработчиков алгоритмов по классификации образов и машинного обучения проблема распознавания рукописных цифр является хорошим примером для тестирования производительности классификации.

Нейронные сети активно используются для решения задач классификации в системах компьютерного зрения. Нейронные сети оптимизированы для раздельного обучения с учителем с целью разделения