

Настоящий блок лекций подготовлен при поддержке «ЦРТ | Группа компаний» и Университета ИТМО





Обработка речевых сигналов Блок 1. Цифровая обработка сигналов

 Цифровой сигнал
 Нейронные сети
 Преобразование Фурье

 Спектр и спектрограмма
 Цифровой фильтр

 Теорема отсчетов
 Частота дискретизации

Владимир Волохов

Научный сотрудник ООО «ЦРТ-инновации», доцент, к.т.н.



Блок 1. Цифровая обработка сигналов



text-snadow. filter: dropshadow(color= color:#777: Theader Amain-navigation ut li www.shadow box-shadows -moz-box-shadow hackground-color #F9F9F9 n ut ti actitus Часть 4. Цифровая обработка сигналов и машинное

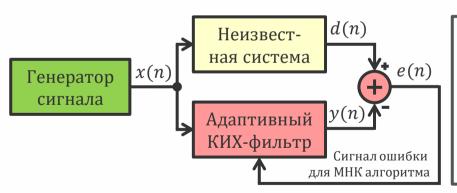
обучение

План лекции

- Немного про адаптивную фильтрацию!
- Нерекурсивные и рекурсивные фильтры это нейронные сети!
- Нерекурсивный фильтр и ResNet-блок
- Банк фильтров и Inception-блок
- «Свёртка» в свёрточных нейронных сетях
- Понижение частоты дискретизации и пулинг
- Рекурсия и фильтрация в методе моментов
- Цифровая обработка сигналов и голосовая биометрия
- Мел-частотные кепстральные коэффициенты в задаче верификации дикторов
- Синтез фильтров и end-to-end обучение системы верификации дикторов

Немного про адаптивную фильтрацию!

- Адаптивная фильтрация сложившееся научно-техническое направление в цифровой обработке сигналов, которое активно развивается с конца 50-х годов 20 века
- Решаемые задачи: системная идентификация, эхокомпенсация в телефонных сетях и т.п.



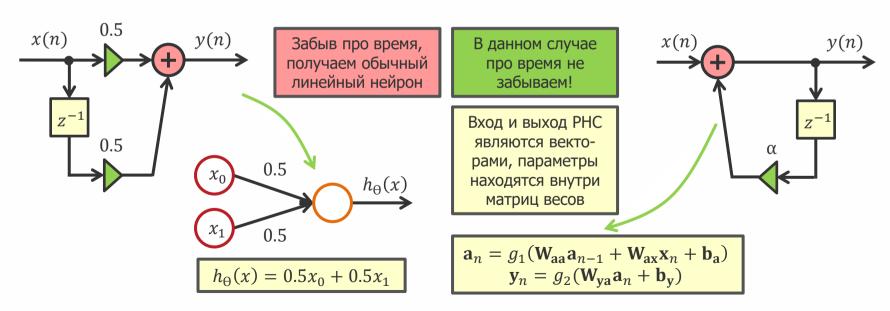
Адаптивный алгоритм наименьших квадратов Уидроу—Хопфа

$$h_{n+1}(k) = h_n(k) + \alpha e(n) x(n-k)$$
 α – параметр сходимости $h(k)$ – коэффициенты КИХ-фильтра $0 \le k \le N-1$ $n=0,1,...$

Адаптивная фильтрация сигналов: теория и алгоритмы [Джиган В.И., 2014]

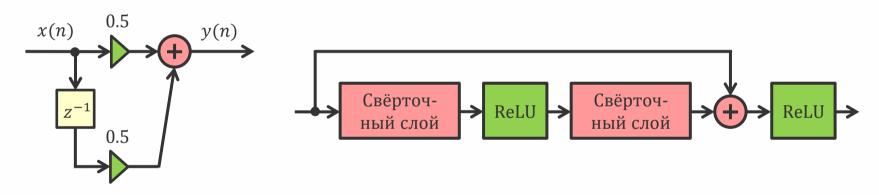


 Нерекурсивные фильтры по своей сути соответствуют нейронным сетям прямого распространения, а рекурсивные фильтры – рекуррентным нейронным сетям (РНС)



Нерекурсивный фильтр и ResNet-блок

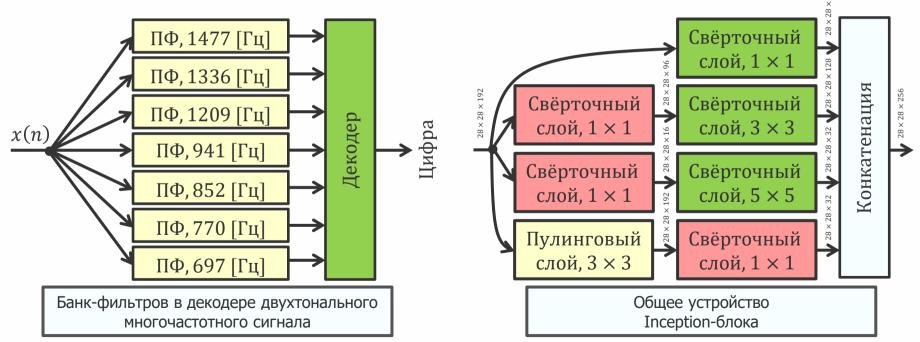
■ В нерекурсивных фильтрах есть обратная связь по входу, как в ResNet-блоке



■ В цифровой обработке сигналов существует отдельная тема, посвященная структурам фильтров, идеи которых могут дать возможность пофантазировать на тему архитектур современных нейронных сетей (Schaum's outlines of digital signal processing [Hayes M.H., 2011])!

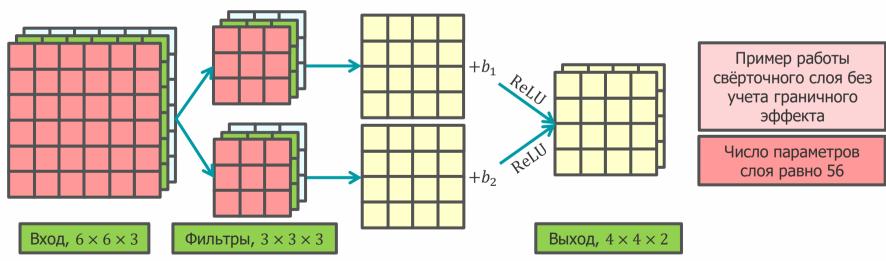
Банк фильтров и Inception-блок

• Банк-фильтров неявно присутствует в Inception-блоке



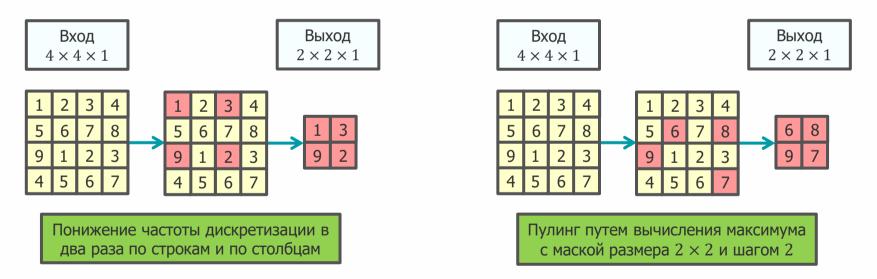
«Свёртка» в свёрточных нейронных сетях

- Внутри свёрточного слоя выполняется не совсем свёртка!
 - При выполнении стандартной свёртки происходит инверсия маски по соответствующим координатам
 - При выполнении стандартной свёртки компонента смещения отсутствует
 - Обычно стандартная свёртка выполняется поканально, а не кросс-канально



Понижение частоты дискретизации и пулинг

 Понижение частоты дискретизации в цифровой обработке сигналов (цифровой обработке изображений) и пулинг со сдвигом маски больше 1 очень похожи по смыслу!



Рекурсия и фильтрация в методе моментов

 Для настройки параметров моделей на практике приходится использовать методы численной оптимизации, например, метод моментов

Градиентный спуск

$$\theta_i^n = \theta_i^{n-1} - \alpha \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_m) \bigg|_{\theta_0^{n-1}, \theta_1^{n-1}, \dots, \theta_m^{n-1}} = \theta_i^{n-1} - \alpha \cdot v_n$$

Метод моментов

$$\theta_i^n = \theta_i^{n-1} - \alpha \cdot \left(\rho \cdot v_{n-1} + \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_m) \bigg|_{\theta_0^{n-1}, \theta_1^{n-1}, \dots, \theta_m^{n-1}} \right), v_0 = 0, \rho \in [0.9, 0.99]$$



• Для настройки параметров моделей на практике приходится использовать методы численной оптимизации, например, метод моментов

Градиентный спуск

$$\theta_i^n = \theta_i^{n-1} - \alpha \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_m) \bigg|_{\theta_0^{n-1}, \theta_1^{n-1}, \dots, \theta_m^{n-1}} = \theta_i^{n-1} - \alpha \cdot v_n$$

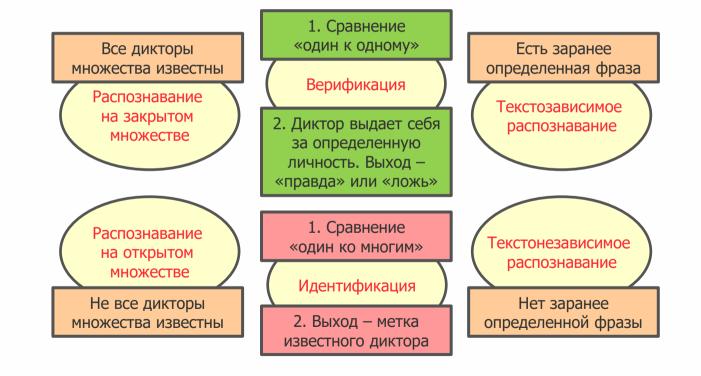
Метод моментов

$$\theta_i^n = \theta_i^{n-1} - \alpha \cdot \left(\rho \cdot v_{n-1} + \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_m) \bigg|_{\theta_0^{n-1}, \theta_1^{n-1}, \dots, \theta_m^{n-1}} \right), v_0 = 0, \rho \in [0.9, 0.99]$$

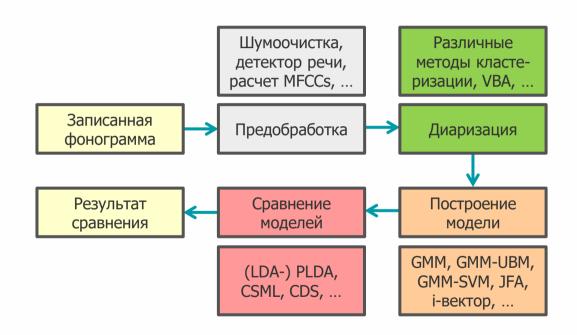
А это, кстати, рекурсивный фильтр!

$$v_n = \rho \cdot v_{n-1} + \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_m) \bigg|_{\theta_0^{n-1}, \theta_1^{n-1}, \dots, \theta_m^{n-1}}$$

Что такое голосовая биометрия?

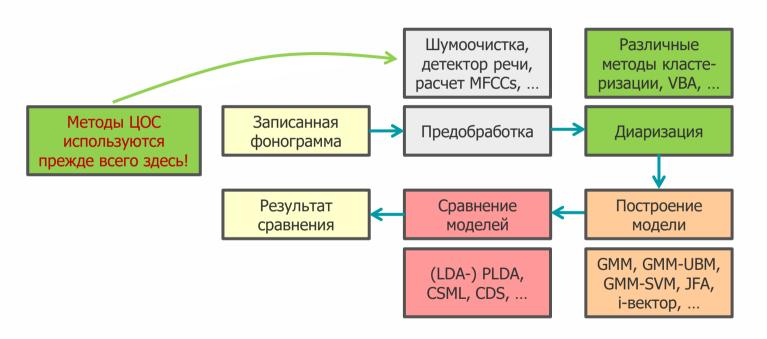


Общая схема системы голосовой биометрии



Speaker recognition by machines and humans [Hansen J.H.L., Hasan T., 2015]

Общая схема системы голосовой биометрии



Speaker recognition by machines and humans [Hansen J.H.L., Hasan T., 2015]

Мел-частотные кепстральные коэффициенты в задаче верификации дикторов

- Модели на основе глубокого обучения, например х-векторные:
 - экстрактор признаков (эмбеддингов) на основе глубокой нейронной сети, обучаемой дискриминативно с использованием стохастического градиентного спуска на базе натурального градиента



• сравнение моделей с использованием подхода на базе линейного дискриминантного анализа и вероятностного линейного дискриминантного анализа

Deep neural network embeddings for text-independent speaker verification [Snyder D., Garcia-Romeo D., Povey D., Khudanpur S., 2017]



Синтез фильтров и end-to-end обучение модели системы верификации дикторов

• Синтез фильтров — процедура автоматического расчета коэффициентов фильтра по некоторым требованиям разработчика

Метод окон

- 1. Задать частотную характеристику идеального фильтра $H_{\text{\tiny M}}(e^{j\widehat{\omega}})$
- 2. Вычислить ИХ идеального фильтра:

$$h_{\mathsf{H}}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{\mathsf{H}}(e^{j\widehat{\omega}}) e^{j\widehat{\omega}n} d\widehat{\omega}, n \in \mathbb{Z}$$

3. Усечь ИХ идеального фильтра окном:

$$h(n) = h_{_{\mathrm{H}}}(n) \cdot w(n)$$
 $w(n) \neq 0, \ 0 \leq n \leq N-1; \ w(n) = 0, \$ иначе

Пример

1. Идеальная частотная характеристика ПФ:

$$H_{\mathsf{M}}(e^{j\widehat{\omega}}) = \begin{cases} 0, & 0 \leq |\widehat{\omega}| < \widehat{\omega}_{\mathsf{H}} \\ e^{-j\alpha\widehat{\omega}}, \widehat{\omega}_{\mathsf{H}} \leq |\widehat{\omega}| \leq \widehat{\omega}_{\mathsf{B}} \\ 0, & \pi \geq |\widehat{\omega}| > \widehat{\omega}_{\mathsf{B}} \end{cases}$$

2. ИХ идеального фильтра ($-\infty < n < \infty$):

$$h_{H}(n) = \frac{\widehat{\omega}_{B}}{\pi} \cdot \operatorname{sinc}(\widehat{\omega}_{B}(n-\alpha)) - \frac{\widehat{\omega}_{H}}{\pi} \cdot \operatorname{sinc}(\widehat{\omega}_{H}(n-\alpha))$$

Speaker recognition from raw waveform with SincNet [Ravanelli M., Bengio Y., 2018]



Синтез фильтров и end-to-end обучение модели системы верификации диктора

- Синтез фильтров процедура автоматического расчета коэффициентов фильтра по некоторым требованиям разработчика
- Набор полосовых фильтров, вычисленных с использованием метода окон для разных частот среза $\widehat{\omega}_{\rm H}$ и $\widehat{\omega}_{\rm B}$, позволяет получить банк фильтров, который можно использовать для обучения признаков «на лету» вместо мел-частотных кепстральных коэффициентов
- Импульсная характеристика каждого фильтра банка будет определяться всего двумя обучаемыми параметрами и может быть достаточно длинной!

$$h(n) = \left(\frac{\widehat{\omega}_{\mathrm{B}}}{\pi} \cdot \mathrm{sinc}(\widehat{\omega}_{\mathrm{B}}(n-\alpha)) - \frac{\widehat{\omega}_{\mathrm{H}}}{\pi} \cdot \mathrm{sinc}(\widehat{\omega}_{\mathrm{H}}(n-\alpha))\right) \cdot w(n)$$
 $w(n) \neq 0, \ 0 \leq n \leq N-1; \ w(n) = 0, \$ иначе; $\alpha = (N-1)/2$

Speaker recognition from raw waveform with SincNet [Ravanelli M., Bengio Y., 2018]

Благодарности

- В лекции использовались материалы следующих книг:
 - Солонина А.И. и др. Основы цифровой обработки сигналов, 2005
 - Prandoni P., Vetterli M. Signal processing for communications, 2008
 - Hayes M.H. Schaum's outlines of digital signal processing, 2011
 - Николенко С., Кадурин А., Архангельская Е. Глубокое обучение, 2018
- В лекции использовались материалы следующих онлайн-курсов:
 - Oppenheim A.V. Digital signal processing (RES.6-008), 1975. Massachusetts Institute of Technology
 - Peters R.A. Image processing (EECE/CS 253), 2007. Vanderbilt University School of Engineering
 - Prandoni P., Vetterli M. Digital signal processing, 2014. Coursera's online course of École polytechnique fédérale de Lausanne
 - Ng A. Convolutional neural networks, 2019. Coursera's online course of Stanford University
 - Ng A. Sequence models, 2019. Coursera's online course of Stanford University
- Отдельное спасибо коллегам по работе, без которых бы эта лекция не состоялась!



Группа компаний ЦРТ

O HAC

В группу компаний ЦРТ входят компании ЦРТ, ЦРТ-инновации и SpeechPro.

ЦРТ – российская компания, разработчик инновационных систем в сфере технологий синтеза и распознавания речи, анализа аудио- и видеоинформации, распознавания лиц, голосовой и мультимодальной биометрии.

ЦРТ-инновации — научно-исследовательская компания, передовой разработчик голосовых и бимодальных биометрических систем. Резидент Фонда «Сколково». Области научно-исследовательской деятельности компании: биометрия по голосу и лицу, распознавание речи, анализ больших данных.

SpeechPro – представительство Группы ЦРТ в Северной Америке с главным офисом в Нью-Йорке. SpeechPro взаимодействует с клиентами и партнерами ЦРТ из США и Канады.

КОНТАКТНАЯ ИНФОРМАЦИЯ

Санкт-Петербург

Адрес: 194044, г. Санкт-Петербург, ул. Гельсингфорсская, 3-11, лит. Д Телефон: (+7 812) 325-88-48

Факс: (+7 812) 327-92-97

Эл. почта: stc-spb@speechpro.com

Москва

Адрес: Москва, ул. Марксистская, д.3, стр.2,

Бизнес-центр «Таганский» Телефон: +7 (495) 669-74-40 Факс: +7 (495) 669-74-44

Эл. почта: stc-msk@speechpro.com