

Низкоскоростное кодирование речевого сигнала с применением радиальных нейронных сетей

Е.А. Лопухова¹, Г.С. Воронков¹, И.В. Кузнецов¹

¹Уфимский государственный авиационный технический университет, Карла Маркса 12, Уфа, Россия, 450008

Аннотация

В статье рассматриваются методы низкоскоростного кодирования на основе нейронных сетей. Анализируется вопрос применения радиальных нейронных сетей для решения задач низкоскоростного кодирования и снижения энергопотребления систем передачи информации без ухудшения качества связи. Описывается построение модели кодека с применением радиальных нейронных сетей в среде MatLab.

Ключевые слова

Низкоскоростное кодирование, энергоэффективные системы связи, радиальные нейронные сети

1. Введение

Для множества современных приложений обработки, хранения и передачи голосовой информации в целях решения проблемы экономии энергетических ресурсов систем связи и снижения требований к пропускной способности канала при потоковой передаче необходимо внесение низкоскоростного кодирования, в частности, кодеров формы волны. Несмотря на то, что кодеры формы волны не могут работать на низких скоростях передачи с приемлемым качеством, подобно параметрическим кодерам, этот недостаток может компенсировать включение в структуру кодеров формы волны нейронных сетей (НС).

Данная работа посвящена построению модели кодека, включающего НС, не требующую высокой вычислительной сложности и снижающую необходимое энергопотребление системы.

2. Низкоскоростное кодирование речевого сигнала с применением нейронных сетей

На сегодняшний день все больше исследований в области низкоскоростного кодирования направлены на создание интеллектуальных систем сжатия речевой информации, основанных на методах машинного обучения, в частности, с включением НС.

Вместе с этим большинство современных решений в этой области включают нейросетевые структуры, требующие высокой вычислительной сложности, что усложняет системы связи и увеличивает потребление энергии. Примерами вышеприведенных исследований могут служить работы [1], [2], [3] и [4].

Наиболее оптимальной структурой для уменьшения динамического диапазона канальных сигналов была определена сеть с одним скрытым слоем. В качестве НС, сжимающей и восстанавливающей динамический диапазон сигнала, была выбрана радиальная НС. Задачу компрессии и декомпрессии сигнала можно свести к аппроксимации сжатой или восстановленной версии входного вектора отсчетов. Конфигурация стандартной радиальной НС представляет собой трехслойную структуру. На входной слой подаются отсчеты входного сигнала, скрытый слой состоит из нейронов с радиальной активационной функцией, выходной слой производит взвешенное суммирование результатов работы скрытого слоя.

Задача аппроксимации функции, заданной обучающей выборкой, представляет собой подбор количества радиальных функций и поиск значений настраиваемых параметров сети.

При N -мерной размерности входного вектора и M -мерном скрытом слое, где $M < N$, решение формируется в виде:

$$F(x) = \sum_{i=1}^M w^i \varphi(\|x - c^i\|) . \quad (1)$$

Моделирование многоканальной системы передачи сообщений было осуществлено в среде MatLab. Для трехканальной системы передачи речевых сообщений были сформированы НС кодера и декодера (Рисунок 1). Скрытый слой включал 27 нейронов. Сформированные сети были обучены на сжатом и исходном речевом сигнале соответственно длительностью 6 с.

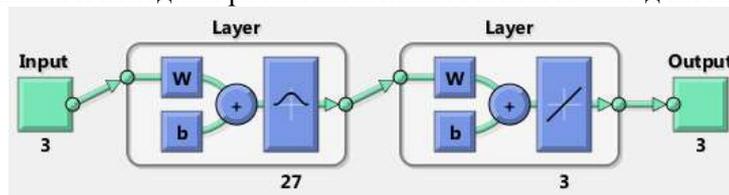


Рисунок 1: Структура нейронных сетей кодера для речевых сигналов

Время обработки сигналов сформированным кодеком составило 11 с, максимальное количество BER= $6,5 \cdot 10^{-3}$. Также были сформированы НС для трехканальной системы передачи музыкальных сообщений длительностью 6 с. Скрытый слой включал 69 нейронов. Изменение значения распространения базисных функций и количества нейронов радиального типа заключалось в отличии динамических свойств речевых и музыкальных сигналов. Время обработки сигналов сформированным кодеком составило 12 с, максимальное значение BER= $7 \cdot 10^{-3}$. Степень сжатия по всем каналам в обоих вычислительных экспериментах составило 70%.

Полученные результаты свидетельствуют о высокой степени сжатия канальных сигналов, обеспечивающей возможное снижение энергопотребления системы передачи при сохранении приемлемого качества речевых сигналов. Основным недостатком рассмотренной модели является высокая задержка кодера при обработке передаваемого сигнала. Для ее снижения предполагается в дальнейшем использовать специализированные вычислительные устройства.

3. Благодарности

Исследование выполнено при поддержке Министерства науки и высшего образования РФ в рамках выполнения работ по Государственному заданию ФГБОУ ВО УГАТУ # FEUE-2020-0007 по теме «Теоретические основы моделирования и семантического анализа процессов преобразования вихревых электромагнитных полей в инфокоммуникационных системах».

4. Литература

- [1] Shlomot, E. Combined harmonic and waveform coding of speech at low bit rates / E. Shlomot, V. Cuperman, A. Gersho // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Seattle, WA, USA. – 1998. – P. 585-588.
- [2] Cernak, M. Phonological vocoding using artificial neural networks / M. Cernak, B. Potard, P. N. Garner // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane. – 2015. – P. 4844-4848.
- [3] Faúndez-Zanuy, M. Nonlinear predictive models computation in ADPCM schemes // 10th European Signal Processing Conference. – 2000. – P. 1-4.
- [4] Yoshimura, T. Speaker-dependent Wavenet-based Delay-free Adpcm Speech Coding / T. Yoshimura, K. Hashimoto, K. Oura, Y. Nankaku, K. Tokuda // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom. – 2019. – P. 7145-7149.
- [5] Слеповичев, И.И. / Курс лекций: Введение в нейроинформатику. – Саратов: СГУ, 2006. – 101 с.