АЛГОРИТМ ОБУЧЕНИЯ МНОГОСЛОЙНОГО НЕЙРОСЕТЕВОГО КВАНТОВАТЕЛЯ АУДИОКОДЕРА

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

Аврамов В.В.

Петровский А.А. – д.т.н., профессор

Сжатие звуковой информации является актуальной задачей в современном мире, поскольку активное развитие и внедрение получают такие технологии как передача аудиоинформации по коммуникационным каналам (VoIP, VoLTE), потоковое вещание мультимедиа (Streaming Media), цифровое радиовещание (DAB). В каждой из перечисленных технологий одной из наиболее важных задач является компактное представление цифрового звукового сигнала. В любом алгоритме сжатия звука важнейшим шагом процесса кодирования является алгоритм квантования данных.

Результатом работы аудиокодера [1] является набор наиболее перцептуально важных для восприятия человеком параметров (атомов). Структура каждого атома представляется как его вес (вещественное число), и два целочисленных параметра характеризующие позицию атома в дереве реконструкции пакетного дискретного вейвлет преобразования декодера. Позиции атомов эффективно кодируются с использованием энтропийного кодирования Хаффмана, в то время как вес каждого атома должен быть квантован для компактного представления и передачи декодеру.

Нейросетевое квантование (NNQ – NeuralNetworkQuantization) представляет собой совместное квантование вектора параметров, представленных вещественными числами, в некоторый дискретный набор. Процесс NNQ устраняет избыточность благодаря эффективному использованию взаимосвязанных свойств векторных параметров. Таким образом, данный подход предлагает несколько уникальных преимуществ по сравнению с скалярным квантованием, включая возможность использования линейных и нелинейных зависимостей между векторными компонентами и гибкость в выборе многомерных форм ячеек квантователя. Определение NNQ формулируется следующим образом: NNQ размера M является отображением $X \in \mathbb{R}^M$ в N-мерный кодовый вектор Y содержащий K дискретных выходных значений.

Исходя из вышеприведенного определения, естественной архитектурой искусственной нейронной сети (ИНС) для реализации квантователя является симметричная сеть прямого распространения (AE - autoencoder), содержащая входной, кодовый и выходной слои. Основной задачей обучения AE является получение на выходе вектора с минимальным отклонением от входного. Описанная однослойная сеть весьма ограниченна по своим вычислительным возможностям, особенно в задаче квантования, предполагающей дискретные ограничения на выходы кодового слоя. Исходя из этого, очевидно, что одного скрытого слоя будет недостаточно, и выбор следует сделать в пользу многослойной архитектуры (DAE – deepautoencoder).

Архитектура DAENNQ была экспериментально выбрана как показано на рисунке 1. Весовые коэффициенты DAE были предварительно обучены тремя однослойными AE: {200×100}, {100×50}, {50×20}, в соответствии со стратегией жадного послойного обучения [2]. Первый AE был обучен на подготовленной тренировочном наборе. Каждый последующий AE был обучен с использованием выходов скрытого слоя предыдущего AE в качестве входных тренировочных данных. Затем, обученные слои были объединены чтобы сформировать DAE, и дальнейшей тонкой подстройки весовых коэффициентов. Параметры модели обновлялись после каждого набора из случайно выбранных 1000 обучающих примеров с использованием алгоритма стохастической оптимизации Adam [3].

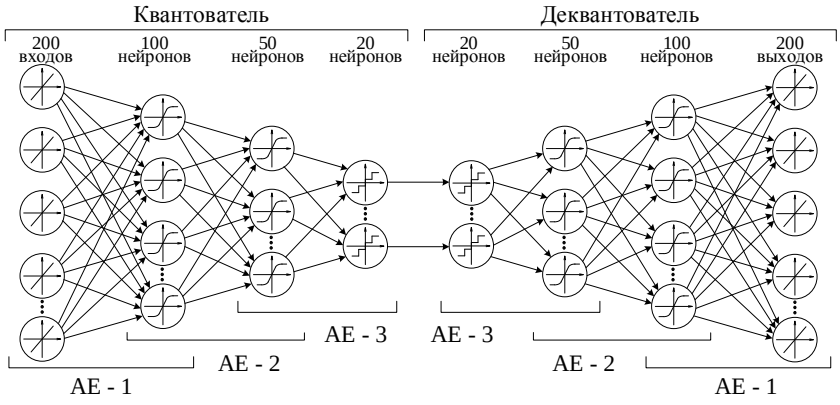

Рис. 1 – Нейросетевой квантователь

Функцией активации первых двух AE является гиперболический тангенс. Для получения дискретного представления в центральном кодовом слое DAE в третьем AE функцией активации выступает ступенчатая функция активации [4], образованная гиперболическими тангенсами. Данная функция, естественным образом, в процессе обучения, отображает входные векторы с вещественными компонентами в дискретный набор координат и описывается следующим выражением:

$$f(z) = \frac{1}{M-1}\sum_{i=1}^{M-1}\left(tanh\left(\alpha\left(z - \frac{2\cdot i}{M} + 1\right)\right)\right), \quad (1)$$

где $\alpha = 100$ – контролирует угол наклона ступеней, $M = 32$ – количество ступеней.

В процессе проведения экспериментов, нейросетевой квантователь был обучен на тренировочной последовательности длительностью около 36 минут, включающей речевые образцы, музыку и другие звуковые сигналы. В качестве тестовой последовательности выступали образцы, описанные в таблице 1.

Табл. 1 – Описание тестовых образцов

Образец	Описание	Образец	Описание
es01	Вокал (Suzan Vega)	si01	Клавесин
es02	Речь на немецком	si02	Кастаньеты
es03	Речь на английском	si03	Камертон
sc01	Соло на трубе	sm01	Волынка
sc02	Оркестр	sm02	Металлофон
sc03	Поп-музыка	sm03	Струнные

Средний бюджет бит для квантованных весов атомов и кодирования положения оценивается следующим образом. Кодовый слой для варианта 200 атомов состоит из 20 нейронов со ступенчатой функцией активации, состоящей из 32 ступеней, что соответствует 5 битам на каждую ступень. Умножение количества нейронов на количество бит на каждый нейрон даст 100-битный бюджет для представления 200 атомов. Таким образом, суммарный средний битрейт составляет около 10 кбит/с, что соответствует эквивалентной степени сжатия около 70. Для экспериментов с другими вариантами числа атомов структура слоев изменяется пропорционально (например, для 300 атомов, кодовый слой состоит из 30 нейронов).

Для оценки качества реконструируемого аудио-сигнала была использована метрика ITU-R Recommendation BS.1387-1 PEAQ (Perceptual Evaluation of Audio Quality). Это метрика используется для оценки перцептуального искажения на основе слуховой модели человеческого уха. Оценка PEAQ- это показатель объективной разности (ODG). Шкала ODG определяется следующим образом: 0 - незаметное ухудшение; -1 - ощутимое, но не раздражающий; -2 - немного раздражает; -3 - раздражает; -4 - очень раздражающее ухудшение. На рисунке 2 показаны результаты оценки качества универсального аудио / речевого кодера в составе с представленным методом квантования.

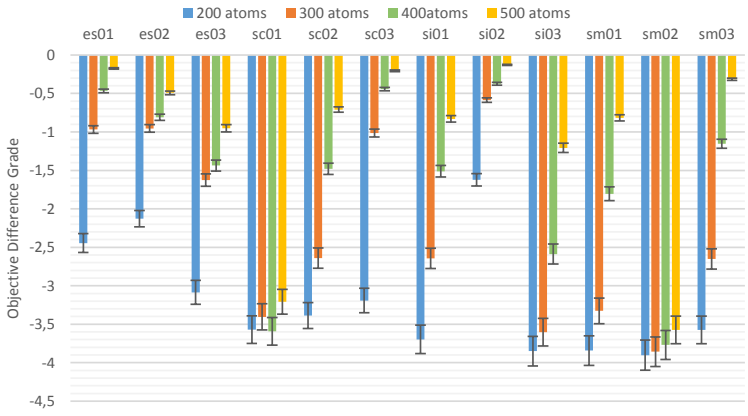

Рис. 2 – Результаты оценки качества реконструированных тестовых образцов

Список использованных источников:

1. Петровский Ал.А., Петровский А.А. Масштабируемые аудиоречевые кодеры на основе адаптивного частотно-временного анализа звуковых сигналов //Труды СПИИРАН, 1(50). – 2017. – с. 56-92.

2. Bengio Y. Popovici, Greedy layerwise training of deep networks / Y. Bengio, P. Lamblin, D. Popovici // NIPS. – 2006. – pp.153–160.

3. Diederik K. Adam: a method for stochastic optimization / K. Diederik, J. Ba // Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015). – 2015.– arXiv preprint: arXiv:1412.6980.

4. Hecht-Nielsen R. Replicator Neural Networks for Universal Optimal Source Coding / R. Hecht-Nielsen // Science, vol. 269, no. 5232. – 1995. – pp. 1860-1863.