**УДК 004.89**

МЕТОДЫ ИЗВЛЕЧЕНИЯ АКУСТИЧЕСКИХ ПРИЗНАКОВ В ЗАДАЧЕ РАСПОЗНАВАНИЯ РЕЧИ РЕКУРРЕНТНЫМИ НЕЙРОННЫМИ СЕТЯМИ С ДОЛГОЙ КРАТКОСРОЧНОЙ ПАМЯТЬЮ

Мошкарова Лилия Айратовна, младший научный сотрудник 1, lmoshkarova@niime.ru, *+7 (915) 480-92-72*

*Тельминов Олег Александрович, к.т.н., начальник лаборатории1,* otelminov@niime.ru, *+7 (495) 229-74-97*

1 АО «НИИМЭ», г. Москва

Аннотация: в работе приведены некоторые методы извлечения признаков из аудиосигнала, решена задача распознавания русской речи с помощью рекуррентной нейронной сети LSTM, проведен анализ результатов распознавания для разных акустических признаков.

Ключевые слова: распознавание русской речи, MFCC, PLP, PNCC, LSTM, CTC.

Введение

Распознавание речи широко применимо в задачах управления системами или голосового поиска. Целью данной работы является определение эффективного способа извлечения признаков из речевого сигнала.

Извлечение акустических признаков

Наиболее популярным методом выделения признаков являются мел-частотные кепстральные коэффициенты (MFCC – mel-frequency cepstrum coefficients [1]). Способ основан на особенностях человеческого слуха и представлении сигнала в мел-шкале. Следующий метод – анализ перцептивного линейного предсказания (PLP – perceptual linear predictive [2]), который связан с неравной восприимчивостью слуха на разных частотах и использованием кривой критической полосы. Менее распространенный, но весьма точный способ – извлечение нормированных по мощности кепстральных коэффициентов (PNCC – power-normalized cepstral coefficients [3]). В методе применяется алгоритм подавления шума, основанный на наблюдении, что мощность речи в канале изменяется быстрее мощности фонового шума.

Распознавание и результаты

Для распознавания речи использовалась двунаправленная рекуррентная нейронная сеть с двумя слоями LSTM – long short-term memory по 128 нейронов. В процессе обучения минимизировалась функция потерь CTC - Connectionist Temporal Classification [4]. Сеть была обучена на наборе данных VoxForge. Ошибка распознавания по показателю LER (Label Error Rate) на полном тестовом наборе составила 27,8% для PNCC, 29,4% для PLP и 29,6% для MFCC. При этом для достижения наилучшего результата для PNCC требуется обучение нейронной сети за меньшее число эпох, чем для других методов. Аппаратная реализация вычислительных устройств сегодня может быть выполнена с помощью устройств традиционной КМОП-технологии [5] типа матричных интегральных схем (БМК или ПЛИС), моделирующих процессорные системы [6], либо специализированных процессоров. При дальнейшем развитии технологий и компонентной базы возможна будет реализация макетов нейронов и сетей на их основе по типу [7], которые позволят реализовать рассматриваемые алгоритмы более эффективно.

Список использованных источников

1. X. Huang, A. Acero, H-W. Hon, Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, Prentice Hall, 2001
2. H. Hermansky. Perceptual linear predictive (PLP) analysis of speech // The Journal of Acoustical Society of America. – 1990. – P. 1738-1752
3. C. Kim, R. M. Stern, Power-normalized cepstral coefficients (PNCC) for robust speech recognition // IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, May 2012
4. A. Graves, S. Fernandez, F. Gomez, J. Schmidhuber, Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural 'networks // [ICML '06: Proceedings of the 23rd international conference on Machine learning](https://dl.acm.org/doi/proceedings/10.1145/1143844). June 2006 – P. 369–376
5. Красников Г.Я., Горнев Е.С., Матюшкин И.В. Общая теория технологии и микроэлектроника: в 3 ч. // Электронная техника. Сер. 3: Микроэлектроника. — Ч. 1. — 2017. — Вып. 1 (165). — С. 59-63; Ч. 2. —2017. — Вып. 4 (168). — С. 16—41; Ч. 3. — 2018. — Вып. 3 (171). — С. 63—93.
6. В.И. Эннс. СнК, БМК или ПЛИС: выбор варианта исполнения цифровой интегральной схемы // Компоненты и технологии. – 2018. – № 4. – C.100–102.
7. Е. С. Горнев, Г. С. Теплов. Математическая модель конечного автомата абстрактного нейрона и сетей на его основе// Нано – и микросистемная техника. –2018. –Т.20. – № 7. –С. 434–442.

METHODS FOR EXTRACTION OF ACOUSTIC CHARACTERISTICS IN THE SPEECH RECOGNITION PROBLEM BY RECURRENT NEURAL NETWORKS WITH LONG SHORT-TERM MEMORY

Lilia Ayratovna Moshkarova, junior researcher1, lmoshkarova@niime.ru, +7 (915) 480-92-72

Oleg Aleksandrovich Telminov, Ph.D., head of the laboratory1, otelminov@niime.ru, +7 (495) 229-74-97

1 JSC MERI, Moscow

Abstract: the paper presents some methods for extracting features from an audio signal, the problem of recognizing Russian speech using a recurrent neural network LSTM are solved, the results of recognition for different acoustic features are analyzed.

Key words: Russian speech recognition, MFCC, PLP, PNCC, LSTM, CTC.