

Преамбула. Решение закрытой задачи

1. Предобработка данных

На рисунке ниже показан процесс предобработки тренировочного набора данных.

В каждом образце была удалена тишина по заданному порогу, который подбирался эмпирически.

После этого, каждый образец разбивался на экземпляры по 2,5 секунды с перекрытием в половину окна.

В качестве характеристик аудио сигнала была выбрана мел-спектограмма, для каждого экземпляра вычислялась мел-спектрограмма с 80 фильтрами.

Экземпляры были аугментированы до наиболее объема наиболее часто встречающегося класса. Общее количество тренировочных примеров равнялось 53416 экземплярам



2. Обучение модели

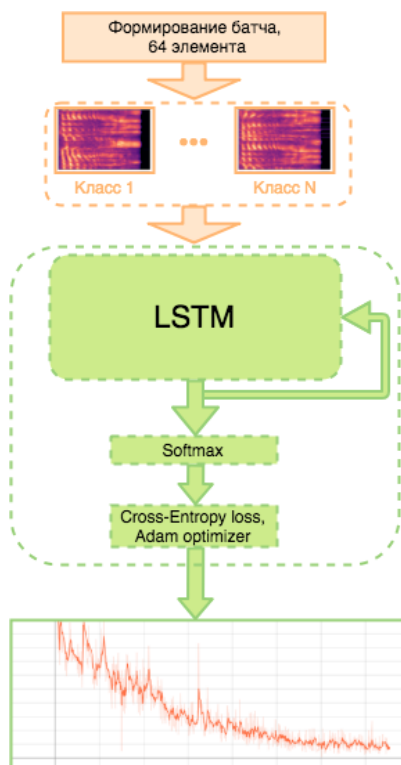
На вход нейронной сети формировались батчи по 64 элемента.

Была выбрана рекуррентная нейронная сеть с простой ячейкой LSTM. Количество входов сети = 80, количество шагов сети = 205, количество скрытых состояний = 30

Выходным слой нейронной сети: Softmax.

В качестве функции потерь использовалась средняя кросс-энтропия.

Нейронная сеть обучалась на протяжении 5 эпох.



Вначале каждой эпохи формировалась валидационная выборка, в качестве метрик были выбраны доля правильных ответов и матрица ошибок.

Точность модели на валидационной выборке составила порядка 93%. Ниже представлена матрица ошибок, полученная в конце обучения.

649	0	5	0	0	7	2	0
0	655	14	1	0	16	0	22
8	6	608	12	36	1	3	1
1	3	5	663	1	0	1	0
1	0	14	0	624	1	1	1
1	2	1	0	0	650	0	96
10	0	8	0	1	2	645	3
0	3	3	0	0	5	1	549

3. Валидация модели

При тестировании модели, для каждого примера удалялась тишина. После этого пример разбивался на образцы, которые подавались в нейронную сеть. Ответом нейронной сети были выходы Softmax слоя для каждого образца, которые суммировались друг с другом и

делились на количество образцов полученных из тестового примера. Максимальное значение из этого вектора и являлось уверенностью классификатора в принадлежности примера к классу.

