ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ТОНАЛЬНОСТИ РЕЧЕВЫХ СИГНАЛОВ ЧЕЛОВЕКА

Талецкий А.И.

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

Калугина М.А. – канд. физ-мат. наук

Одно из наиболее популярных направлений в современном машинном обучении – методы классификации тональности различных данных, таких как фотографии лиц, тексты постов в интернете или записи речевых сигналов человека. Оценка тональности речевых сигналов отстает от других направлений и, в связи с этим, представляет особый интерес для различных исследований.

Данная работа является логическим продолжением и развитием предыдущей. Где для решения задачи тестировались методы классического машинного обучения. Лучший результат среди классических методов показала композиция градиентного бустинга и ближайших соседей 0.71 (или 71%).

В работе используется датасет, загруженный с сайта стэндфордского университета и распространяемый по свободной некоммерческой лицензии. Датасет стоит из коротких аудиофайлов, каждый по 4 секунды. 12 мужчин и 12 женщин зачитывают 2 фразы в 8 тональных классах:

- нейтральность;
- спокойствие;
- счастье;
- грусть, печаль;
- злость, ненависть;
- испуг;
- удивление.

В результате варьирования различных параметров, не влияющих на общую тональность записи, получается датасет содержащий 2800 примеров.

Для оценки качества работы алгоритмов используется оценка на отложенных данных. Для тестирования стратифицировано по тональным классам откладывается 20% или 560 примеров. Подсчет производится по следующим метрике accuracy.

Нейронная сеть состоит из слоев, собранных в блоки в следующем порядке:

- AveragePolling1D;
- GRU [1];
- Dense and Dropout.

Сеть имеет порядка 300 тысяч параметров. В качестве оптимизатора используется Adam [2]. Одна эпоха обучения с бачом в 64 объекта занимает 45 секунд. Уже к 10 эпохе сеть преодолевает бейзлайн в 71% и достигает максимума в 97% примерно к 25 эпохе. Эффект переобучения практически отсутствует. На тестовых данных сеть показывает результат в 0.8857 или примерно 89%.

Как и предполагалось рекуррентные нейронные сети оказались весьма эффективными для решения этой задачи. Прирост качества составил почти 20% что превышает все возможные ожидания.

В дополнение стоит отметить, что дополнительных улучшений можно достигнуть путем использования слоя Attention. Также имеет смысл попробовать одномерные сверточные сети.

Список использованных источников:

1.Cho, Kyunghyun; van Merrienboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation".

2.Diederik P. Kingma; Jimmy Ba; (2014). “Adam: A Method for Stochastic Optimization”.