**Определение эмоционального состояния диктора по голосу с использованием многослойной нейронной сети**

***Сабо Кристина Вячеславовна***

*Студент*

*Алтайский государственный университет,*

*институт цифровых технологий, электроники и физики, Барнаул, Россия*

*E-mail: sabo.kristina@yandex.ru*

Задача автоматического распознавания эмоциональной окрашенности звучащей речи находит применение во множестве областей: разработке роботов, автономных транспортных средств, оборудовании для нейро-обратной связи, полиграфии, защите информации, частной собственности. Необходимость учета действия внешних факторов, влияющих на характеристики голоса и вносящих искажения в регистрируемый речевой сигнал, обуславливает необходимость разработки новых методов определения эмоционального состояния.

В данной работе для решения задачи распознавания эмоционального состояния диктора был использован стандартный подход, предполагающий сочетание экстрактора признаков и классификатора. В качестве значимых информативных признаков речевого сигнала использовались мел-частотные кепстральные коэффициенты [1]. Классифицирующий алгоритм представлял собой нейронную сеть [2], состоящую из трех слоев (входного, скрытого и выходного). Для сжатия параметров модели и предотвращения переобучения использовалась L2-регуляризаци [3]. В качестве функции потерь применялась логистическая функция потерь. Обучение сети осуществлялось методом стохастического градиентного спуска с моментами.

При обучении и оценке качества предложенной модели использовались данные из набора RAVDESS [4]. Он делился в отношении 3:1 на обучающую и тестовые выборки. В число различных эмоциональных состояний входили: спокойствие, счастье, грусть, зависть, испуг, удивление, отвращение и нейтральная эмоция. Образцы эмоционально окрашенной речи данного набора записывались в студии профессиональными актерами.

Проведенный численный эксперимент показал, что выбранный подход хорошо работает для бинарной классификации при различении диаметрально противоположных эмоций. Оцениваемое на тестовом наборе качество составляло 96%. Дальнейшее развитие данной работы будет заключаться в выборе более подходящей архитектуры нейронной сети для эффективной классификации по голосу диктора множества промежуточных эмоциональных состояний.

**Литература**

1. Сидоров, К.В. К вопросу оценки эмоциональности естественной и синтезированной речи по объективным признакам / К.В. Сидоров, М.В. Калюжный // Вестник Тверского государственного технического университета. Вып. 18. Тверь, 2011. С. 81−85.
2. Изилов, Ю.Я Многослойная персептронная нейронная сеть в задаче моделирования речевых сигналов / Ю.Я. Изилов // Информационно – управляющие системы – 2003 – №2-3. – С. 46-50.
3. Каширина, И.Л. Исследование и сравнительный анализ методов оптимизации, используемых при обучении нейронных сетей / И.Л. Каширина // Системный анализ информационных технологий – 2018 - №4. – С. 123 – 132.
4. Livingstone S. R., Russo F. A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English / S. R. Livingstone // PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.