

МЕТОДЫ АУГМЕНТАЦИИ АУДИОДАНЫХ

Лукьянов Павел Александрович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: lukyanovpavel1998@gmail.com

Научный руководитель — *Дьяконов Александр Геннадьевич*

Понятию аугментации сложно дать точное определение, в данной работе под аугментацией понимается создание новых данных с помощью модификации уже имеющихся. Использование аугментации может быть особенно полезно для небольшой обучающей выборки и может улучшить обобщающую способность модели, являясь мощным инструментом в борьбе с переобучением.

В данной работе предложен метод аугментации аудиоданных SwapVerticalStripes, основанный на перестановке вертикальных полос в мел-спектрограмме. Для исследования применимости предложенного метода в задаче классификации вычислительные эксперименты проведены с использованием двух датасетов: Heartbeat Sounds [2] (звуки сердцебиения) и GTZAN [3] (классификация музыкальных жанров). Метрика качества — процент верно классифицированных объектов. В рамках экспериментов использовались модели нейронных сетей resnet18 [4] и resnet50 [4]. Результаты экспериментов, представленные в таблицах 1 2, показали возможную применимость предложенного метода в задаче аудиоклассификации.

На практике выбирается некоторый набор заранее заданных методов аугментации. RandAugment [1] — одна из наиболее популярных стратегий использования этих методов. В данной работе предложен алгоритм применения методов аугментации аудиоданных с выбором конкретного метода аугментации после каждой эпохи обучения. Результаты экспериментов, представленные в таблицах 3 4, показали возможное преимущество предложенного алгоритма над RandAugment [1].

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	81.98 ± 2.34	82.23 ± 2.4
SwapVerticalStripes	83.2 ± 1.3	83.65 ± 1.07

Таблица 1: Результаты экспериментов (Heartbeat Sounds [2]) с предложенным методом аугментации SwapVerticalStripes

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	74.3 ± 3.03	73.0 ± 3.24
SwapVerticalStripes	76.6 ± 2.67	75.6 ± 3.68

Таблица 2: Результаты экспериментов (GTZAN [3]) с предлагаемым методом аугментации SwapVerticalStripes

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	81.98 ± 2.34	82.23 ± 2.4
RandAugment [1]	83.1 ± 0.92	84.57 ± 1.3
Предлагаемый алгоритм	86.65 ± 0.67	86.75 ± 0.76

Таблица 3: Результаты экспериментов (Heartbeat Sounds [2]) с предлагаемым алгоритмом применения методов аугментации

Метод аугментации	resnet18	resnet50
Аугментация отсутствует	74.3 ± 3.03	73.0 ± 3.24
RandAugment [1]	75.0 ± 2.61	74.9 ± 2.63
Предлагаемый алгоритм	76.8 ± 1.75	72.2 ± 2.8

Таблица 4: Результаты экспериментов (GTZAN [3]) с предлагаемым алгоритмом применения методов аугментации

Литература

1. Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, Quoc V. Le. RandAugment: Practical automated data augmentation with a reduced search space // arXiv preprint arXiv:1909.13719. — 2019.
2. Bentley, P. and Nordehn, G. and Coimbra, M. and Mannor, S. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. — 2011.
<http://www.peterjbentley.com/heartchallenge/index.html>
3. G. Tzanetakis and P. Cook. Musical genre classification of audio signals // IEEE Transactions on Speech and Audio Processing. — 2002.
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition // In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.