

ОБУЧЕНИЕ ПРЕДСТАВЛЕНИЙ БИОАКУСТИЧЕСКИХ СИГНАЛОВ МЕТОДАМИ БЕЗ УЧИТЕЛЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ

Симонова Александра Денисовна

Студент

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

E-mail: s02230504@gse.cs.msu.ru

Научный руководитель — Кравцова Ольга Анатольевна

Автоматический анализ биоакустических сигналов является важной задачей в области экологического мониторинга и изучения поведения животных. Однако применение современных методов машинного обучения в данной области существенно ограничено нехваткой размеченных данных, получение которых требует значительных экспертных усилий. В связи с этим особый интерес представляют методы обучения представлений без учителя (self-supervised learning, SSL), позволяющие извлекать информативные признаки непосредственно из неразмеченных акустических сигналов.

Недавние исследования показали, что модели, обученные без учителя на человеческой речи, могут эффективно использоваться для анализа биоакустических сигналов и обеспечивать конкурентоспособные результаты в задачах классификации [1, 2]. Такие модели, как HuBERT [3], Wav2Vec 2.0 [4] и WavLM [5], формируют акустические представления на основе различных стратегий обучения, включая masked prediction, контрастивное обучение и задачи подавления шума. Несмотря на это, влияние типа self-supervised задачи на качество представлений в задачах биоакустики остаётся недостаточно изученным.

В данной работе исследуется обучение представлений биоакустических сигналов методами без учителя для решения задачи классификации. В отличие от существующих подходов, фокусирующихся на сравнении доменов предобучения, основное внимание уделяется анализу свойств сформированных SSL-представлений и их пригодности для биоакустических задач. В качестве примера рассматриваются вокализации обыкновенных мармозеток (*Callithrix jacchus*), обладающие сложной акустической структурой.

Предлагаемый подход основан на использовании предварительно обученных акустических моделей, реализующих различные стратегии обучения. Извлечённые представления анализируются с точки зрения их информативности, устойчивости к шуму и способности

к переносу между различными биоакустическими датасетами. Для этого проводится оценка представлений, извлекаемых из различных уровней моделей, а также их применение в задачах классификации типов сигналов и идентификации источников звука с использованием лёгких классификационных надстроек.

Ожидается, что проведённое исследование позволит:

оценить влияние типа self-supervised обучения на структуру акустических представлений;

выявить уровни моделей, наиболее информативные для биоакустических задач;

определить устойчивость различных представлений к акустическим и доменным вариациям.

Таким образом, работа направлена на исследование и сравнительный анализ self-supervised представлений для биоакустических сигналов и повышение эффективности их применения в условиях ограниченной разметки.

Литература

1. Sarkar E. Can Self-Supervised Neural Representations Pre-Trained on Human Speech distinguish Animal Callers? // In Proc. of Interspeech. 2023.
2. Sarkar E. Comparing Self-Supervised Learning Models Pre-Trained on Human Speech and Animal Vocalizations for Bioacoustics Processing. // arXiv:2501.05987. 2025.
3. Hsu W. HuBERT: Self-supervised speech representation learning by masked prediction of hidden units // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2021. vol. 29. pp. 3451–3460.
4. Baevski A. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // Advances in Neural Information Processing Systems. 2020. vol. 33. pp. 12 449–12 460.
5. Chen S. WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing // IEEE Journal of Selected Topics in Signal Processing. 2022. vol. 16. no. 6.