ОПТИМИЗАЦИЯ МЕТОДА ДЕТЕКТИРОВАНИЯ ВМЕШАТЕЛЬСТВА И ЗАЩИТЫ БИОМЕТРИЧЕСКИХ ДАННЫХ В АУДИОПОТОКЕ

Вернигорова А.А. (Университет ИТМО), Талынкова Е.Н. (Университет ИТМО), Костин А.А. (Университет ИТМО)
Научный руководитель – Роговой В.
(Университет ИТМО)

Введение:
Сейчас интерес к использованию голосовой биометрии растет, но она уязвима для атак с фальсификацией голоса. Развитие технологий создает возможность создания качественных дипфейков, которые могут обойти системы автоматической верификации по голосу. Это может привести к серьезным последствиям, включая несанкционированный доступ к банковским счетам. Существующие методы защиты не всегда эффективны, поэтому нужны новые меры противодействия фальсификации голоса, которые могут быть надежно использованы для защиты систем ASV.

Основная часть:
Для решения проблемы безопасности систем автоматической верификации предлагается применение методов машинного обучения, в частности, разработка и использование нейронной сети с архитектурой LSTM для обнаружения аудио-спуфинга. Архитектура LSTM способна выявлять долгосрочные зависимости, что крайне важно для аудиозаписей, так как позволяет анализировать характеристики на протяжении всей аудиозаписи, а не только на отдельных участках. Входными данными для нейронной сети является аудиозапись, представленная в виде временного ряда в массиве numpy. Из аудиозаписи извлекаются различные акустические характеристики (например, спектральные и частотные характеристики, коэффициенты кепстрального преобразования и т.д.).
Мы выделяли три категории диференцированночастотных файлов, при этом основной упор строился на анализе файлов с частотой звука не превышающей 7кГц. Этот диапазон частот обычно содержит важные акустические особенности, которые могут быть значимы для идентификации дипфейков. Анализ аудиозаписей в этой категории помогает улучшить обобщение модели на разнообразные типы аудиосигналов. Категории высоко- и назкочастотных аудиофайлов выделялись по выделенным критериям, таким как частота звука и длинна волны.
Мы проводили очистку от фонового шума, применяя различные методы фильтрации и подавления шума, такие как фильтры низких и высоких частот, а также алгоритмы шумоподавления, включая методы временного окна и методы подавления шума на основе спектрального профиля. Это позволило обеспечить чистоту аудиозаписей, которые послужили основой для анализа и обучения модели.
После очистки мы проводили отбор наиболее важных акустических признаков, используя анализ подтвержденных чистых и синтезированных записей. Мы применили матрицу корреляции для определения признаков, сильно влияющих на различение подлинных записей от дипфейков, и выбрали признаки с высокой корреляцией с целевой переменной для дальнейшего использования в обучении нейронной сети.
Это позволяет оптимизировать решение и уменьшить ошибки первого и второго рода. На выходе нейронная сеть предоставляет классификацию аудиозаписи (дипфейк или подлинная). Таким образом, разрабатываемый метод обладает высокой точностью в определении принадлежности аудиозаписи к классу дипфейков или подлинных записей и выявлении характерных изменений в акустических характеристиках, связанных с дипфейками.

Вывод:

В рамках данного исследования были рассмотрены различные методы распознавания фальсификации голосовых шаблонов, используемых для обмана систем автоматической верификации говорящего. Анализируя недостатки этих методов, нами был разработан собственный подход к обнаружению вмешательства в аудиопоток с целью неправомерного использования биометрических данных. Наш метод позволяет повысить точность определения аудио-спуфинга и снизить количество ошибок первого и второго рода путем уменьшения компонент при анализе аудиофайла и использования временных рядов.

Полученные результаты имеют потенциал для широкого применения в различных областях, где используются системы автоматической верификации говорящего, включая сферу финансовых операций. Наш подход поможет повысить уровень безопасности и предотвратить несанкционированное использование биометрических данных при проведении финансовых транзакций.

Список использованных источников:

1. Awais Khan, Khalid Mahmood Malik, James Ryan1, and Mikul Saravanan. Voice Spoofing Countermeasures: Taxonomy, State-of-the-art, experimental analysis of generalizability, open challenges, and the way forward. arXiv:2210.00417v2 [eess.AS] 21 Nov 2022.
2. Chadha A, Abdullah A, Angeline L, Sivanesan S (2021) A review on state-of-the-art Automatic Speaker verification system from spoofing and anti-spoofing perspective. Indian Journal of Science and Technology 14(40): 3026-3050.
3. Aakshi Mittal, Mohit Dua. Automatic speaker verification systems and spoof detection techniques: review and analysis. International Journal of Speech Technology (2022) 25:105–134.
4. Xuechen Liu, Xin Wang, Md Sahidullah, Jose Patino, Héctor Delgado, Tomi Kinnunen, Massimiliano Todisco, Junichi Yamagishi, Nicholas Evans, Andreas Nautsch, Kong Aik Lee. ASVspoof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild. arXiv:2210.02437v1 [cs.SD] 5 Oct 2022.
5. Junichi Yamagishi, Xin Wang, Massimiliano Todisco, Md Sahidullah, Jose Patino, Andreas Nautsch, Xuechen Liu, Kong Aik Lee, Tomi Kinnunen, Nicholas Evans, H´ector Delgado. ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection. arXiv:2109.00537v1 [eess.AS] 1 Sep 2021.
6. R. Yang, J. Chen, Q. Jin, "Deep Learning for Audio-Visual Emotion Recognition: Recent Advances and Future Directions", IEEE Transactions on Affective Computing, 2020
7. T. Kinnunen, Z. Wu, E. Sizov, "Voice Spoofing Countermeasures: Recent Advances and Future Directions", IEEE Signal Processing Letters, 2019