

0.1. Попов Д.Н. Дистилляция ансамбля для псевдоразметки в диаризации дикторов *зе А. И.*

Диаризация дикторов — задача определения временных границ речевых сегментов и их принадлежности дикторам, число которых неизвестно [1]. Точное разделение речи важно для распознавания речи и суммаризации, особенно в сложных акустических условиях, таких как многоязычные встречи или шумные разговоры.

Задача диаризации осложняется пересекающейся речью, шумом, разнообразием акустических условий и неизвестным числом дикторов. Традиционные методы (кластеризация, спектральный анализ) неустойчивы в реальных сценариях и требуют дополнительной обработки [1]. Нейронные подходы, такие как EEND [2], точнее, но зависят от больших объемов размеченных данных [3–5]. Ручная разметка трудоемка и плохо масштабируема, особенно для специфичных данных. Синтетические данные [6] не всегда отражают реальность, снижая генерализацию моделей. Методы обучения без учителя и с частичным обучением с учителем ограничены отсутствием надежных псевдометок.

Псевдоразметка широко применяется в смежных областях, таких как обработка естественного языка и распознавание речи [7]. В диаризации дикторов псевдоразметка страдает от накопления ошибок и отсутствия учета неуверенности предсказаний. Кроме того, до сих пор не были предложены методы ансамблирования нескольких диаризационных алгоритмов с оценкой неуверенности для повышения надежности псевдометок [8].

В работе представлен новый метод псевдоразметки, основанный на взвешенном ансамблировании выходов нескольких диаризационных алгоритмов и способ оценки неуверенности ансамбля моделей. Метод тестировался на датасетах, характеризующихся сложными акустическими условиями, с использованием модели EEND. Эксперименты проводились в сценариях с ограниченными размеченными данными и большими неразмеченными корпусами.

Модели, обученные с использованием предложенного метода, достигли уровня диаризационной ошибки, сопоставимого с обучением на размеченных данных. При добавлении неразмеченных данных метод показал превосходство над подходами, использующими только размеченные данные, что подчеркивает его масштабируемость в условиях ограниченных ресурсов.

Работа выполнена при поддержке научно-образовательной школы МГУ "Мозг, когнитивные системы, искусственный интеллект" и НИР по госбюджетной теме МГУ 5.1.21 "Вероятностные модели глубинного обучения, процедуры их настройки и применения при решении прикладных задач анализа данных".

Научный руководитель — к.ф.-м.н. Майсурад-

Список литературы

- [1] PARK T. J. ET AL. A review of speaker diarization: Recent advances with deep learning // //Computer Speech & Language. – 2022. – Т. 72. – С. 101317.
- [2] FUJITA Y., KANDA N., HORIGUCHI S., XUE Y., NAGAMATSU K., WATANABE S. End-to-end neural speaker diarization with self-attention // 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). 2019. P. 296–303.
- [3] FUJITA Y., KOMATSU T., SCHEIBLER R., KIDA Y., OGAWA T. Neural diarization with non-autoregressive intermediate attractors // ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2023. P. 1–5.
- [4] HAN E., LEE C., STOLCKE A. BW-EDA-EEND: streaming end-to-end neural speaker diarization for a variable number of speakers // ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. P. 7193–7197.
- [5] LIU Y., HAN E., LEE C., STOLCKE A. End-to-end neural diarization: from transformer to conformer // arXiv preprint arXiv:2106.07167. 2021.
- [6] EDWARDS E., BRENNDOERFER M., ROBINSON A., SADOUGH N., FINLEY G. P., KORENEVSKY M., AXTMANN N., MILLER M., SUENDERMANN-OEFT D. A free synthetic corpus for speaker diarization research // Speech and Computer. 2018. P. 113–122.
- [7] HIGUCHI Y., MORITZ N., LE ROUX J., HORI T. Momentum pseudo-labeling for semi-supervised speech recognition // arXiv preprint arXiv:2106.08922. 2021.
- [8] STOLCKE A., YOSHIOKA T. DOVER: A method for combining diarization outputs // 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). – IEEE, 2019. – С. 757–763.