УДК 004.934.5

МЕТОД ЛАТЕНТНОЙ ДИФФУЗИИ ДЛЯ ПОСТРОЕНИЯ ГЛАДКОГО СЕМАНТИЧЕСКОГО ПРИЗНАКОВОГО ПРОСТРАНСТВА В ДИСКРЕТНОЙ МОДЕЛИ ЭМОЦИЙ

Свищев А. Н. (Университет ИТМО)
Научный руководитель – д.т.н. Матвеев Ю.Н. (Университет ИТМО)

В данной работе рассматривается подход к построению латентного семантического пространства акустических признаков с помощью метода латентной диффузии.

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР №622281 «Разработка методов и алгоритмов для мультимодального распознавания валентности высказываний и доминантности дикторов в полилогах».

Введение. Построение признаковых акустических пространств для речевого сигнала является важным этапом как в генеративных, так и в дискриминантных задачах. Гладкость и семантический характер получаемых признаковых пространств играют ключевую роль для интерпретируемости результатов при классификации и для управления результатом при задаче генерации.

Основная часть. Наиболее значимые результаты в области построения гладких семантических пространств для речевого сигнала были достигнуты в подходе wav2vec [1]. Однако такое пространство имеет очень большую размерность и сложно поддается метрический и семантической интерпретации. Последние успехи в области генерации изображений и аудио с помощью метода латентной диффузии [2] показывают его эффективность в части получения гладких обусловленных семантических признаковых пространств. В работе исследуется применение такого подхода для построения обусловленного на дискретные эмоциональные категории данных латентного диффузного пространства на примере корпуса dusha [3].
Выводы. В ходе работы был проведен эксперимент по обучению двух моделей. Исследовано полученное признаковое пространство на предмет семантической и метрической гладкости. Сделаны выводы о преимуществах и недостатках нового подхода, а также возможным сфер применения полученных результатов.

Список использованных источников:

1. Baevski A. et al. wav2vec 2.0: A framework for self-supervised learning of speech representations //Advances in neural information processing systems. – 2020. – Т. 33. – С. 12449-12460.
2. Rombach R. et al. High-resolution image synthesis with latent diffusion models //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2022. – С. 10684-10695
3. Бимодальный эмоциональный корпус dusha. Режим доступа: https://github.com/salute-developers/golos/tree/master/dusha#dusha-dataset (дата обращения: 06.03.2023).