

УДК 004.8

РАЗРАБОТКА ЭКСТРАКТОРА ГОЛОСОВЫХ ДЕСКРИПТОРОВ НА ОСНОВЕ НЕЯВНЫХ СЛОЕВ ДЛЯ ЗАДАЧИ ВЕРИФИКАЦИИ ДИКТОРА

Корневская А.М. (ИТМО)

Научный руководитель – кандидат технических наук Новоселов С.А.
(ИТМО)

Введение. Как правило, нейронные сети представляют собой композицию функций, каждая из которых явным образом зависит от значения, поступающего на вход. Слои, которые задаются совместным условием на вход и выход слоя, не всегда могут быть представлены явным образом как функция выхода от входа. Тем не менее, заданные таким образом слои при определенных условиях могут быть интегрированы в нейронные сети. В представленной работе разрабатывается экстрактор голосовых дескрипторов на основе современных архитектур с применением двух различных видов неявных слоев: нейронные дифференциальные уравнения [1] и DEQ-слои (от англ. Deep Equilibrium Models [2]).

Основная часть. Современные архитектуры верификации диктора (ResNet34 [3], wav2vec 2.0. [4]) модифицируются для возможности интеграции неявных слоев в обучение на задачу дикторской верификации. Полученные системы обучаются на базе VoxCeleb2 [5] и тестируются на протоколе VoxCeleb1-O. Проводится сравнительный анализ построенных систем с классическими вариантами архитектур для задачи верификации диктора по следующим параметрам:

- 1) качество верификации в терминах равновероятной ошибки;
- 2) количество параметров;
- 3) скорость работы.

В сравнении участвуют системы, построенные на базе двух видов неявных слоев: нейронные дифференциальные уравнения и DEQ. Применение DEQ-слоев позволило снизить равновероятную ошибку на протоколе тестирования по сравнению с классическими архитектурами без увеличения количества параметров.

Выводы. Разработан экстрактор голосовых дескрипторов на основе неявных слоев. Проведено сравнение полученных архитектур с современными системами. Выявлены положительные и отрицательные стороны такого вида обучения для задачи дикторской верификации. Применение технологии DEQ обеспечило прирост качества верификации на протоколе VoxCeleb1-O.

Список использованных источников:

1. Chen R. T. Q. et al. Neural ordinary differential equations //Advances in neural information processing systems. – 2018. – Т. 31.
2. Bai S., Kolter J. Z., Koltun V. Deep equilibrium models //Advances in neural information processing systems. – 2019. – Т. 32.
3. Koonce B., Koonce B. ResNet 34 //Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. – 2021. – С. 51-61.
4. 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.
5. Chung, J.S., Nagrani, A., Zisserman, A. VoxCeleb2: Deep Speaker Recognition // Proc. Interspeech. – 2018. – С. 1086-1090, doi: 10.21437/Interspeech.2018-1929.