

УДК 004.932.2

## ПРИМЕНЕНИЕ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ ПРИ КРОСС-КОРПУСНОМ И КРОСС-КУЛЬТУРНОМ НЕПРЕРЫВНОМ РАСПОЗНАВАНИИ ЭМОЦИЙ

Дресвянский Д. В.

(Университет ИТМО)

Научный руководитель – д.т.н., доцент, Карпов А.А.

(Санкт-Петербургский институт информатики и автоматизации Российской академии наук)

В данной работе рассмотрено применение глубокого обучения при кросс-корпусном и кросс-культурном непрерывном распознавании эмоций. Представлено описание архитектуры глубокой нейронной сети и данных. Проведен сравнительный анализ результатов исследования и современных подходов.

В современном мире информационные технологии и автоматизация приобретают все большую важность в жизни человека. Бурное развитие машинного обучения позволяет решать ранее недоступные машине проблемы. Одной из таких задач стала задача распознавания эмоций. В последнее время проведено немало исследований в области распознавания эмоций. Однако, большинство исследований фокусировались на одной конкретной базе данных, что приводит к кросс-корпусной и кросс-культурной проблемам – неспособности модели выйти за рамки распределения базы данных, на которой она была обучена.

В настоящее время в области распознавания эмоций доминируют две модели – категориальная модель (эмоции представлены заданным количеством категорий) и модель Дж. Рассела (эмоции характеризуются знаком и интенсивностью, которые принимают значения от -1 до 1). В данной работе используется вторая модель, позволяющая отображать весь эмоциональный спектр. Основным инструментом при распознавании эмоций стали глубокие нейронные сети (ГНН). Однако эффективность ГНН зависит от выбранной архитектуры и имеющихся данных.

В данном исследовании предлагается следующая архитектура ГНН:

- Предобученная сверточная нейронная сеть (СНН) VGGFace2 – извлечение глубоких признаков из изображения;
- Три полносвязных слоя – редукция извлеченных признаков;
- Три слоя долгой краткосрочной памяти (LSTM) – выявление и захват временных закономерностей в данных.

Одним из способов преодоления кросс-корпусной и кросс-культурной проблем является обучение модели на нескольких стандартизированных базах данных (БД) с различными национальностями и языками участников. Стандартизация в данном случае означает приведение всех БД к одному временному шагу при аннотировании, а также соответствующую предобработку каждого фрейма.

В данной работе использовались следующие БД: AffectNet, AFEW-VA, RECOLA, SEMAINE, Aff-Wild и SEWA. AffectNet, задействованная для настройки второй части ГНН (три полносвязных слоя), представляет собой набор из 440 тысяч изображений, размеченных в терминах модели Дж. Рассела (знак и интенсивность эмоции). Остальные БД, представляющие собой видеоряды реакций и диалогов различных людей, были стандартизированы и применены для обучения третьей части модели.

Результаты данного исследования могут быть использованы для преодоления кросс-корпусной и кросс-культурной проблем и построения универсальной системы распознавания эмоций, которая требуется повсеместно – от робототехники и медицины до промышленных предприятий.