

УДК 004.89

АДАПТАЦИЯ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ ДЛЯ РАБОТЫ С НАБОРАМИ ДАННЫХ ДЛЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

Вампилов Б.А. (Университет ИТМО)

Научный руководитель – доцент, кандидат технических наук Забашта А.С.
(Университет ИТМО)

Введение. Генеративно-состязательные модели относительно недавно стали применяться для генерации синтетических данных [1] путем обучения на множестве настоящих выборок и создания новых синтетических выборок. В работе представляются предварительные результаты по созданию моделей генерации табличных наборов данных и их классификации в контексте задач мета-обучения и автоматического машинного обучения, основанных на принципах, развиваемых в методах генеративно-состязательных сетей.

Основная часть. Табличные наборы данных обладают свойством инварианта, когда таблицы остаются неизменными при перестановке строк или столбцов, которое ограничивает использование сверточных нейронных сетей. Подход с применением сверточных нейронных сетей был применен в LM-GAN [2], где инвариант обучаемой функции был реализован с помощью стабилизации таблицы, когда, по тому или иному принципу, столбцы и строки таблицы сортировались и в последующем подавались на вход модели. Такой подход имеет недостатки: оператор свертки обрабатывает итеративно только небольшую область таблицы и предполагает, что в признаковом пространстве задан порядок, в то время как в табличных признаках его нет. Разработанная архитектура призвана исправить этот подход, с помощью использования DeepSets [3]. Для DeepSets которой описаны обучаемые инвариантная и эквивариантная функции, агрегирующие значения, что позволяет использовать их в задаче мета-классификации для получения скрытых мета-признаков, аналогично получению обычных мета-признаков из набора данных. Такие функции позволяют построить архитектуру нейронной сети, которая представляет табличные наборы данных в виде множества множеств [4], на которых не задан порядок элементов, и, позволяющая, таким образом, предобрабатывать таблицы перед их подачей на сверточные слои дискриминатора и мета-классификатора, и соответственно позволяющая эффективно обучить генератор.

Выводы. Была разработана архитектура и обучена модель GAN, на основе DeepSets, генерирующая табличные наборы данных, а также мета-классификатор, решающий задачу мета-обучения, и проанализированы полученные результаты относительно существующих решений.

Список использованных источников:

1. Bourou S. et al. A review of tabular data synthesis using GANs on an IDS dataset //Information. – 2021. – Т. 12. – №. 09. – С. 375.
2. Kachalsky I. et al. Generating datasets for classification task and predicting best classifiers with conditional generative adversarial networks //Proceedings of the 3rd International Conference on Advances in Artificial Intelligence. – 2019. – С. 97-101.
3. Zaheer M. et al. Deep sets //Advances in neural information processing systems. – 2017. – Т. 30.
4. Гарипов Р.И. Разработка архитектуры глубокого обучения способной обрабатывать табличные данные для решения задач мета-обучения //Выпускная квалификационная работа. 2023