

МЕТОД ФОРМИРОВАНИЯ ПРОСТРАНСТВА ПАЙПЛАЙНОВ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ МЕТАДАННЫХ

Алиев М.Р. (Университет ИТМО)

Научный руководитель – к.т.н., доцент Муравьев С.Б.
(Университет ИТМО)

Введение. Автоматическое проектирование пайплайнов машинного обучения (МО) для задачи обучения с учителем решается различными оптимизационными методами [1] и это влечет за собой высокую вычислительную сложность процесса. Однако в то же время существует альтернатива, заключающаяся в использовании метаданных. Среди прочего, особенно популярен подход на основе метаобучения [2]. Одним из наиболее современных его представителей является auto-sklearn [3], продемонстрировавшее высокий уровень результатов в сравнении с существующими подходами, добившись значительного преимущества относительно времени работы. В предложенном автором данного материала решении представлено развитие выдвинутой ранее идеи по проектированию пайплайнов МО на основе логического вывода из базы знаний [4], но уже с применимостью к произвольному набору данных. В частности, из основного: будут исследованы различные типы пространств метаданных, будет предложено развитие метода формирования пространства пайплайнов МО и будет проведено экспериментальное сравнение с auto-sklearn.

Основная часть. Для того, чтобы успешно решить поставленную задачу были запланированы следующие шаги:

- 1 Исследование различных типов метаданных с целью выбора типа для собственной разработки, среди которых: показатели качества интерпретируемых и легковесных алгоритмов МО, статистические и информационно-теоретические показатели, характеристики дерева поддержки принятия решений, сформированного на основе набора.
- 2 Сбор данных, среди которых: сущности предметной области (алгоритмы и их гиперпараметры, метаданные, метрики и т.д.), реализация пайплайнов МО, наборы данных для предварительного наполнения базы знаний.
- 3 Разработка метода формирования пространства пайплайнов МО. Предлагается для этого использовать предварительно накопленные знания из графовой базы и вычислять сходство между метаданными, например, на основе корреляции Пирсона [5].
- 4 Программная реализация предложенного метода и механизма непрерывного расширения базы знаний.
- 5 Экспериментальное сравнение качества сформированных пространств пайплайнов с auto-sklearn.

Выводы. Касаемо ожидаемых исследовательских результатов — это добиться более менее стабильных рекомендаций удовлетворительного уровня в сравнении с auto-sklearn. Касаемо реализации - это программный пакет, позволяющий любому желающему спроектировать пайплайн МО под заданные набор данных и задачу обучения с учителем.

Список использованных источников:

- 1 Escalante H. Automated Machine Learning--a brief review at the end of the early years. [Электронный ресурс]. - Режим доступа: <https://arxiv.org/abs/2008.08516>, свободный. Яз. англ. (дата обращения 05.03.2023).
- 2 Vanschoren J. Meta-Learning: a Survey. [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/1810.03548>, свободный. Яз. англ. (дата обращения 05.03.2023).
- 3 Feurer M., et al. Auto-sklearn: Efficient and Robust Automated Machine Learning // Automated Machine Learning. 2019. P. 113–134.
- 4 Алиев М.Р. Метод автоматического наполнения онтологии метаобучения //

Сборник трудов XI Конгресса молодых ученых. 2022. Т. 1. С. 10-13.

5 Cohen I., et al. Pearson correlation coefficient // Noise reduction in speech processing. 2009. P. 1-4.