

УДК 004.8

**АЛГОРИТМ ПО ОТБОРУ ПРИЗНАКОВ С УЛУЧШЕННОЙ СТАБИЛЬНОСТЬЮ
ДЛЯ НАБОРОВ ДАННЫХ С МАЛЫМ КОЛИЧЕСТВОМ ОБЪЕКТОВ**

Пильненьский Н.Д. (Университет ИТМО)

Научный руководитель – кандидат технических наук, Сметанников И.Б.
(Университет ИТМО)

Настоящий доклад рассматривает алгоритм отбора признаков с улучшением стабильности для наборов с малым количеством данных. Предлагается использовать продвинутое метрики ранжирования и баланса классов для улучшения качества моделей машинного обучения.

В настоящее время существует довольно большой спектр задач, содержащий количество признаков на порядок больше, чем число объектов в выборке. В таких наборах данных задача отбора признаков является ключевой для построения моделей с приемлемой точностью. Существует множество решений данной проблемы, такие как: отбор признаков основываясь на значениях статистических функций, отбора на основе качества работы модели, отбор основываясь на результатах работы модели. Также существуют модели использующие различные гибридные подходы, которые представляют суперпозицию указанных ранее. Проблема выбора признаков для набора данных с малым числом объектов является одной из наиболее сложных. Отбираемые для таких наборов данных значимые признаки могут достаточно сильно варьироваться в зависимости от того, как проводилось разбиение на подвыборки при валидации, что именуется низкой стабильностью и сигнализирует о низкой достоверности выбранных признаков.

Предлагаемое решение использует ассемблирование различных методов и ранжирование для повышения стабильности отобранного набора признаков.

Предложенный метод имеет потенциал стать новым state-of-the-art методом для валидации и отбора признаков на наборах данных с малым количеством объектов. Результаты продемонстрированы на выборках генетических данных.

Пильненьский Н.Д. (автор)

Подпись

Сметанников И.Б. (научный руководитель)

Подпись