22. Многоэкземплярное обучение на данных с неполной и неточной разметкой

г.н.с., д.т.н. В.Б. Бериков, н.с., к.т.н. О.А. Кутненко,
лаборатория машинного обучения и анализа данных (И2)

Разработан метод многоэкземплярного обучения на данных с неполной и неточной разметкой, основанный на поиске информативных признаков, фильтрации выборки и применении коллективного решающего правила. Предложено обобщение функции конкурентного сходства (FRiS-функции) на случай задачи многоэкземплярного обучения.

Разметка данных для построения моделей машинного обучения является трудоемким процессом и часто требует привлечения значительных ресурсов, особенно в случае большого объема информации. Поэтому для данных с неполной и неточной разметкой актуальна разработка эффективных методов машинного обучения, называемых методами слабо-контролируемого обучения (Weakly Supervised Learning, WSL).

Во многих практических задачах метка может быть присвоена не отдельному объекту, а целой группе объектов (“пакету”, “bag”), например, отсчетам временного интервала, конформациям молекулы, набору сегментов изображения. Такой вариант постановки задачи WSL – многоэкземплярное обучение (Multiple Instance Learning, MIL) – рассмотрен в работе [1]. Разработан метод решения, основанный на поиске информативных признаков, фильтрации выборки и применении коллективного решающего правила. Предложено обобщение функции конкурентного сходства (FRiS-функции) на случай задачи MIL. FRiS-функция позволяет определить меру сходства пакета с одним образом в конкуренции с другим. В процессе фильтрации удаляются пакеты обучающей выборки, для которых мера сходства со своим образом ниже заданного порога. Экспериментальное исследование эффективности метода проведено в задаче идентификации белков. Выявлены преимущества предложенного алгоритма в сравнении с рядом существующих аналогов.

Работа выполнена в рамках государственного задания ИМ СО РАН, проект FWNF-2022-0015, а также при частичной поддержке гранта ГФ NAP23490529 МНВО Республики Казахстан.

Результат опубликован в работе:

1. *В.Б. Бериков, О.А. Кутненко.* Многоэкземплярное обучение со слабой разметкой на основе выбора информативных признаков и фильтрации выборки // Вычислительные технологии. 2026. Т. 31. №1. С. 92-105. DOI: 10.25743/ICT.2026.31.1.008