

## АНАЛИЗ МЕТОДОВ НАХОЖДЕНИЯ ВЫБРОСОВ В ОБУЧАЮЩИХ ВЫБОРКАХ<sup>1</sup>

*к.т.н, доц. Е.В. Волченко, магистр И.Ю. Кузьменко, Институт информатики и искусственного интеллекта Донецкого национального технического университета, г. Донецк*

Задача нахождения выбросов (цензурирования) объектов выборок является одной из центральных задач интеллектуального анализа данных и решается на этапе предобработки исходных данных.

В общем случае выбросом принято называть объект некоторого класса, значения признаков которого существенно отличаются от значений признаков объектов этого же класса. Такие объекты в пространстве признаков находятся в окружении объектов другого класса и, как следствие, ухудшают качество решающих правил классификации [1]. Наиболее существенной данная проблема является для адаптивных систем, характеризующихся наличием возможности добавления новых данных в процессе работы, что приводит к необходимости постоянного анализа выборок с целью их сокращения и цензурирования. Существующие методы нахождения выбросов можно разделить на две группы. Методы первой группы направлены на корректировку выбросов, что позволяет сохранить размер выборки, но может приводить к ошибкам из-за неверной корректировки. Методы второй группы удаляют выбросы путем фильтрации данных, что может приводить к их уменьшению на 10-15% [2]. Большинство фильтров оценивают данные по выбранной мере сходства объектов некоторого класса на основе гипотезы локальной компактности. Примерами таких фильтров являются алгоритмы STOLP, FRiS-STOLP [2], алгоритмы бустинга и баггинга [1].

В [3] предложен переход к взвешенным обучающим выборкам, каждый  $w$ -объект которых строится по множеству объектов исходной выборки. Введенный дополнительный параметр  $w$ -объектов вес позволяет хранить информацию о числе заменяемых объектов и их взаиморасположении. Поэтому построение фильтров на основе анализа веса  $w$ -объектов является, на наш взгляд, перспективным направлением в решении задачи нахождения выбросов в данных.

**Список литературы:** 1. *Борисова И.А.* Сходство и компактность / *И.А. Борисова, В.В. Дюбанов, Н.Г. Загоруйко, О.А. Кутненко* // Доклады 14-й Всероссийской конференции "Математические методы распознавания образов (ММРО-14)". – М.: Макс-Пресс, 2009. – С. 89 – 92. 2. *Дюличева Ю.Ю.* О задачах фильтрации обучающих данных / *Ю.Ю. Дюличева* // Искусственный интеллект. – Донецк: ИПИИ, 2006. – №2. – С. 65 – 71. 3. *Волченко Е.В.*

---

<sup>1</sup> Работа выполнена при содействии гранта Президента Украины для поддержки научных исследований молодых ученых №GP/F32/130