

ОТЗЫВ

на автореферат диссертационной работы Сарина Константина Сергеевича «Методология построения интерпретируемых нечетких классификаторов, основанных на правилах», представленной на соискание ученой степени доктора технических наук по специальности 2.3.8 – Информатика и информационные процессы

Возрастающие мощности современных вычислительных устройств способствовали развитию систем искусственного интеллекта, основанных на методах машинного обучения. Построенные с помощью таких методов прогностические модели не используют знания экспертов, и в большинстве своем даже разработчики систем не могут объяснить, почему модель пришла к тому или иному решению. Поэтому применение таких систем вызывает недоверие к полученному результату и опасения по поводу причинения вреда. О необходимости объяснения причин принятия решений системами искусственного интеллекта отмечается в «Национальной стратегии развития искусственного интеллекта на период до 2030г.», а также национальном стандарте РФ ГОСТ Р 59276-2020 «Системы искусственного интеллекта. Способы обеспечения доверия».

Нечеткие классификаторы способны предоставить аргументы в пользу принятого решения. Основанные на продукционных правилах с лингвистическими терминами и принятии решений с помощью вывода нечеткой логики они представляют обоснование результата в вполне понятной для человека форме на естественном языке. Современные методы построения, основанные на анализе данных, имеют сложности с построением интерпретируемых классификаторов. Полученные модели трудно перевести в понятное для человека объяснение, как самого решения, так и закономерности, моделируемой классификатором. Таким образом, диссертационное исследование Сарина К.С. направлено на решение актуальной научной проблемы построения интерпретируемых нечетких классификаторов для систем объяснимого искусственного интеллекта.

В диссертации предлагается методология построения нечетких классификаторов, которая основана на концептуальных положениях, устраняющих недостатки современных методов. Недостатки выявлены в ходе аналитического обзора и связаны с низким уровнем интерпретируемости полученных моделей. Методология предполагает применение следующих методов: 1) многокритериальных алгоритмов оптимизации, для построения классификаторов с соблюдением компромисса «точность–интерпретируемость», 2) методов разделения данных на обучающую и валидационную выборку, 3) методов оптимизации, основанных на производных, для параметрической оптимизации классификаторов.

В ходе исследования были получены результаты, которые имеют научную новизну. Разработана методология построения интерпретируемых нечетких классификаторов, определяющая методы анализа данных и их взаимодействие в ходе построения. Разработан алгоритм смешанной многокритериальной оптимизации «кукушкин поиск» с генетическим оператором скрещивания. Разработан дискретный алгоритм оптимизации на основе распределения вероятностей. Адаптирован алгоритм градиентного спуска для параметрической оптимизации нечетких классификаторов. Разработан метод разделения данных на обучающую и валидационную выборки с применением бинарных алгоритмов оптимизации. На основе методологии, с применением указанных выше алгоритмов и метода разделения предложена методика построения нечетких классификаторов основанных на правилах.

Высокая обоснованность и достоверность научных результатов подтверждается следующим: 1) глубоким аналитическим обзором современных методов построения нечетких классификаторов, основанных на правилах; 2) корректным применением методов оптимизации, нечеткой логики, машинного обучения и математической статистики; 3) проведением обширных экспериментов предложенных решений с помощью стандартных

