

О т з ы в

на автореферат диссертации Забежайло М.И.

«Комбинаторные средства формализации эмпирической индукции», представленной на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 – Теоретические основы информатики

Задача обучения на прецедентах хорошо известна специалистам в области компьютерного анализа данных в первую очередь в постановке, предусматривающей обработку числовых данных в рамках тех или иных метрических моделей. Однако, в целом ряде современных приложений computer science актуальной оказывается потребность в математических моделях методов и алгоритмах, позволяющих в процессе обучения на прецедентах иметь дело с *нечисловыми* объектами, особенности описания *структуры* которых оказываются существенными при анализе и прогнозе их свойств. Актуальные примеры таких задач даёт, в частности, область компьютерного прогнозирования полезных (лекарственных) или же, наоборот, – опасных свойств (канцерогенности, мутагенности, токсичности и т.п.) физиологически активных химических соединений. Не менее важная область приложений этого же типа – прогнозирование наличия углеводородного сырья на тех или иных территориях по определенным комплексам структурно-геологических признаков и некоторым специальным комбинациям значений физических величин, замеряемых в процессе электромагнитного зондирования соответствующих территорий. Задачи подобного класса очень часто классифицируют как задачи Big Data, подчеркивая необходимость находить зависимости и принимать решения в процессе таким образом организованного анализа эмпирических данных, в котором компьютерная система обеспечивает *усиление интеллектуальных возможностей* (термин, введенный еще У.Р. Эшби) эксперта. Критически важным элементом организации человеко-машинного взаимодействия в таких исследовательских ситуациях оказывается необходимость построения формализованных компьютерных моделей рассуждений специалиста, в том числе – корректная формализация используемых им эвристик.

В своей диссертационной работе М.И. Забежайло рассматривает задачу формализации эмпирической индукции (формализации комплекса эвристик – индуктивных рассуждений в стиле Д.С. Милля, рассуждений по аналогии в стиле Д. Пойа и абдуктивных объяснений в стиле Ч.С. Пирса) как задачу обучения на прецедентах, которые описываются в виде нечисловых объектов – множеств признаков, связанных определенными отношениями (множеств, графов, цепочек символов, ...), внутренняя структура которых полностью характеризует особенности присущих этим объектам свойств.

Для анализа данных и извлечения скрытых в них эмпирических зависимостей используется алгебраическая формализация сходства объектов (прецедентов) как бинарной операции. Такой подход позволяет выделять структурные «носители» («причины» и «контрпричины») изучаемых свойств как те или иные определенные фрагменты описаний анализируемых прецедентов, порождаемые в процессе вычисления структурных сходств – результатов последовательного применения операции сходства к подмножествам исходных прецедентов. Далее в процессе порождения соответствующих классов толерантности (классов сходства) и реконструкции покрытия этих классов подмножествами прецедентов имеющейся обучающей выборки (классами эквивалентности, порождаемыми всеми общими подструктурами – результатами вычисления операции сходства на подмножествах прецедентов исходной выборки) формируются специальные комбинаторные объекты – диаграммы вложимости классов эквивалентности, каждый из которых базируется на одном из всех возможных для данной обучающей выборки результате вычисления операции сходства. Внутренняя структура пар таких диаграмм, построенных отдельно на исходно заданном множестве примеров и

множестве контрпримеров, полностью определяет возможности прогноза свойств нового объекта путем анализа переносимости на него эмпирических зависимостей, которые порождаются из имеющейся обучающей выборки.

В рамках развиваемого подхода автором диссертационной работы предложен механизм порождения математически корректного ответа на вопрос о возможности прогнозирования свойств каждого нового (заданного своим структурным описанием) прецедента переносом на него эмпирических зависимостей, которые могут быть извлечены из обучающей выборки. Развита техника анализа данных распространена на расширения структурных описаний объектов векторами числовых значений существенных параметров (например, на описания химических структур, дополненных числовыми данными о дозах, концентрациях, температурах и т.п.). Построены оценки вычислительной сложности переборных задач формирования диаграмм вложимости классов эквивалентности прецедентов. Доказана (причем как для структурных, так и для числовых компонентов описаний анализируемых прецедентов) принадлежность ряда важных задач этого типа к классам трудно-разрешимых (NP- и перечислительно полных) комбинаторных проблем.

Для управления комбинаторным перебором, который возникает при проверке переносимости порождаемых эмпирических зависимостей с исходного множества объектов на новые прецеденты, разработан комплекс моделей и алгоритмов, реализующих метод последовательных приближений при порождении диаграмм взаимной вложимости классов эквивалентности (так называемый приближенный ДСМ-метод). Показана возможность реализации этого метода в режиме параллельных вычислений.

Практическая значимость предложенных математических моделей и алгоритмов подтверждена результатами, полученными при решении прикладных задач в нескольких предметных областях.

Полученные автором в рамках диссертационного исследования результаты опубликованы в 37 работах (в том числе – в 17 публикациях в изданиях из перечня ВАК).

Анализ автореферата позволяет заключить, что рассматриваемая диссертация «Комбинаторные средства формализации эмпирической индукции» полностью соответствует требованиям Положения о порядке присуждения ученых степеней ВАК Минобрнауки РФ, а ее автор – М.И. Забежайло – заслуживает присуждения ему ученой степени доктора физико-математических наук по специальности 05.13.17 – Теоретические основы информатики.

Заместитель руководителя Центра оптико-нейронных технологий
Федерального государственного учреждения

«Федеральный научный центр

Научно-исследовательский институт системных исследований

Российской академии наук (ФГУ ФНЦ НИИСИ РАН)»

доктор физико-математических наук

 (В.Г. Редько)

08 июня 2015 года

Подпись д.ф.-м.н. В.Г. Редько удостоверяю

Заместитель директора по научной и методической работе
ФГУ ФНЦ НИИСИ РАН, к.ф.-м.н.

 (А.А. Прилипко)

Редько Владимир Георгиевич, ФГУ ФНЦ НИИСИ РАН,
117218, Москва, Нахимовский просп., 36, к.1, +7 (499) 135-13-51, vgredko@gmail.com