

Отзыв

на автореферат диссертации М.И.Забежайло
«Комбинаторные средства формализации эмпирической индукции»
на соискание ученой степени д.ф.-м.н. по специальности
05.13.17 – «Теоретические основы информатики»

Практическое применение ДСМ-метода можно рассматривать как технологию интеллектуального анализа данных, использующую формализованные правдоподобные рассуждения по индукции, аналогии и абдукции. С помощью формализованной индукции получаются гипотезы о возможных причинах интересующих пользователя свойств объектов (целевых свойств). С помощью рассуждений, которые разработчики ДСМ-метода называют рассуждениями по аналогии, получают гипотезы о наличии или отсутствии целевых свойств. Рассуждения по индукции и аналогии могут выполняться неоднократно. Завершается применение ДСМ-метода формальным аналогом абдукции Ч.С.Пирса, с помощью которой проверяется корректность набора исходных данных и соответствия этим данным применяемых методов формализованных рассуждений.

В терминологии, принятой в сообществе специалистов, занимающихся ДСМ-методом, гипотезы о возможных причинах наличия/отсутствия свойств принято называть гипотезами первого рода, а гипотезы просто о наличии/отсутствии свойств – гипотезами второго рода. С.О.Кузнецов ввел для этих понятий более простые и наглядные термины. Гипотезы о возможных причинах он называет просто *гипотезами*, а гипотезы о наличии/отсутствии свойств – *предсказаниями*.

Уже достаточно длинная история применения ДСМ-метода показывает, что этот метод может давать хорошие результаты для относительно небольших наборов сложно структурированных данных. Табличное представление таких данных может включать большое количество (несколько тысяч) столбцов (признаков) и сравнительно небольшое количество (несколько десятков или сот) строк (объектов). Однако при увеличении количества объектов ДСМ-метод может столкнуться с проблемой резкого увеличения количества гипотез.

Задача сужения множества порождаемых гипотез для ДСМ-метода всегда была актуальной и очень привлекательной. В идеале такое множество должно обладать той же выразительной силой, что и множество всех гипо-

тез, порождаемых обычным ДСМ-методом, т.е. позволять сделать все те предсказания наличия или отсутствия целевых свойств у объектов, которые позволяет сделать множество гипотез, порождаемых обычным ДСМ-методом. Кроме того, желательно, чтобы это множество гипотез было минимальным по включению множеством, позволяющим сделать упомянутую выше совокупность предсказаний.

Задача минимизации множества порождаемых ДСМ-методом гипотез при сохранении их выразительной силы является весьма непростой. Даже частичное ее решение, например, просто сокращение количества таких гипотез при сохранении или при незначительном уменьшении выразительной силы может рассматриваться как большой успех в развитии теории ДСМ-метода, позволяющий реализовать прикладные программы, решающие такие задачи, которые обычный ДСМ-метод не может решить из-за комбинаторного взрыва.

М.И.Забежайло предлагает «приближенный ДСМ-метод», который можно рассматривать как один из вариантов решения проблемы минимизации множества гипотез. Следует также подчеркнуть, что предлагаемые в диссертации алгоритмы допускают эффективное распараллеливание, что для ДСМ-метода, имеющего высокую вычислительную сложность, является весьма полезным качеством.

Приближенный ДСМ-метод основан на оригинальных алгоритмах, предложенных автором диссертации. Основная идея – сократить множества примеров (объектов) и элементарных структурных единиц, составляющих объекты, (признаков), оставив только неразложимые в соответствующей решетке элементы. Кроме того, приближенный ДСМ-метод позволяет не порождать все возможные гипотезы, а ограничиться порождением только тех гипотез, которые удовлетворяют некоторому заранее заданному критерию. Диссертант называет такие гипотезы «полезными». В работе рассматриваются различные примеры таких критериев.

Применение приближенного ДСМ-метода позволило бы расширить область применения этой технологии интеллектуального анализа данных, позволило бы использовать ее и в случае совокупности данных большого объема, что в настоящее время является особенно актуальным в связи с развитием такого направления исследований как Big Data.

Недостатки работы:

- автор часто использует оригинальную терминологию и символику, хотя есть стандартные аналоги (это затрудняет понимание работы),
- автореферат излишне многословен, а диссертация имеет слишком большой объем,
- в автореферате имеются ошибочные утверждения, например:
 - используемое в Утверждении 2.1.1 отношение эквивалентности на самом деле отношением эквивалентности не является,
 - Теорема 5.2.3 верна не всегда, а лишь в том случае, когда решетка замыканий Галуа является дистрибутивной,
- для ряда утверждений из раздела 4 (4.2.10, 4.2.20 и др.) имеются более общие утверждения, доказанные в работах С.О.Кузнецова, можно было бы не приводить их независимых доказательств, а получить их как следствия из известных результатов.

Однако следует отметить, что от упомянутых выше некорректных утверждений основные результаты диссертации не зависят. Развиваемый в работе приближенный ДСМ-метод, его формальное математическое описание и алгоритмы являются серьезным вкладом в то направление интеллектуального анализа данных, которое называется ДСМ-методом автоматического порождения гипотез.

Профессор кафедры математики,
логики и интеллектуальных систем
в гуманитарной сфере
Российского государственного
гуманитарного университета,
д.ф.-м.н., проф.

О.М.Аншаков

