

ОТЗЫВ

**официального оппонента о диссертации Михаила Ивановича Забежайло
«Комбинаторные средства формализации эмпирической индукции»,
представленной на соискание ученой степени доктора физико-
математических наук по специальности 05.13.17. -Теоретические основы
информатики.**

Анализ экспериментальных данных является составной частью исследовательского арсенала практически всех естественно-научных дисциплин. Среди соответствующих методов выделяются методы *интеллектуального* анализа данных, особенностью которых является использование логических и семантических отношений на элементах и структурах данных. В таких случаях требуется привлечение логических средств, в частности, методов логической индукции. Применение индуктивных методов не требует репрезентативных выборок данных, что позволяет использовать основанные на них техники в таких областях, где получение статистически значимых выборок примеров или прецедентов нежелательно или просто невозможно. Рецензируемая работа посвящена исследованию именно таких методов, что свидетельствует о её высокой актуальности.

В работе рассматриваются методы эмпирической индукции, реализованные в рамках известного ДСМ-метода, и основанных на нем программных средств. Указанный метод продемонстрировал свою эффективность для решения некоторых классов задач интеллектуального анализа данных. Как уже было сказано, к числу достоинств ДСМ-метода следует отнести его применимость для анализа статистически незначимых наборов данных, для которых статистические методы слабо применимы. Однако обратной стороной этого достоинства является то обстоятельство, что использование указанного метода для анализа сравнительно большой совокупности данных может приводить к так называемому «комбинаторному взрыву», а именно к порождению такого множества гипотез, что для их анализа требуется коллектив экспертов. Таким образом, возникает задача минимизации множества порождаемых гипотез при сохранении их объяснительных возможностей.

В работе М.И.Забежайло весьма детально рассмотрены принципы индуктивного порождения гипотез, метод сходства, метод различий, метод сходства – различий. Проанализированы основные схемы индуктивного вывода, рассмотрены ДСМ-метод и используемые в нем эвристики и стратегии. Однако, основные результаты диссертации М.И.Забежайло в области теоретических основ информатики связаны с разработанным

автором методом дискретной оптимизации перебора вариантов и созданием на этой основе «приближенного ДСМ–метода». Суть метода состоит в следующем: предлагается начинать с полиномиально быстрого порождения такого подмножества множества всех глобальных сходств структур прецедентов, которое позволяет выполнить все возможные доопределения новых объектов из некоторого исходно заданного множества. Затем, усложняя вычисления, в процессе управляемой и целенаправленной навигации в диаграмме вложений сходств последовательно строить все более и более точные приближенные описания этой диаграммы.

Для обоснования этой техники показано, что каждая диаграмма может быть представлена объединением поддиаграмм специального вида - так называемых псевдо-деревьев (в корне каждого из которых находится описание одного из исходно заданных прецедентов, а остальные вершины содержат лишь входящие в корень элементы - образующие).

Предложен метод порождения приближенных псевдо-деревьев, первым шагом которой является порождение так называемого каркаса – диаграммы взаимной вложимости замыканий Галуа всех одноэлементных подмножеств находящегося в корне псевдо-дерева описания прецедента.

Для этой цели выделены быстро (полиномиально) проверяемые необходимые и достаточные условия, при выполнении которых соответствующий фрагмент псевдо-дерева представляет собою полный гиперкуб на некотором подмножестве образующих.

Для случая, когда фрагмент псевдо-дерева составляет лишь некоторую (требующую восстановления) часть такого гиперкуба, разработан метод последовательных приближений, а именно целенаправленного порождения все более сложных описаний архитектурных фрагментов подобного типа.

Предложенный метод позволяет порождать в первую очередь те фрагменты восстанавливаемых диаграмм, которые полезны для прогнозирования свойств новых прецедентов.

Дано строгое обоснование корректности предложенного автором метода последовательных приближений.

Показано, что этот метод позволяет организовать процесс порождения целевых диаграмм вложимости в режиме параллельных вычислений.

Далее автором рассматриваются вопросы устойчивости порождаемых эмпирических зависимостей при расширении множества прецедентов. Показана функциональность используемого в ДСМ–методе отношения причинности.

Текст снабжен достаточным числом примеров, иллюстрирующих основные идеи. В последней главе приведен ряд прикладных задач, решаемых с помощью предложенной автором техники.

В этой связи можно выделить следующие основные научные результаты работы:

1. Идентифицирован и исследован специальный класс комбинаторных объектов - диаграмм взаимной вложимости различных типов описаний объектов-прецедентов.

2. Процедурная конструкция ДСМ-метода расширена на новые типы данных (структурные описания нечисловых объектов, дополненные числовыми значениями существенных параметров).

3. Предложен метод дискретной оптимизации и управления комбинаторным перебором вариантов при формировании диаграмм вложимости и анализе переносимости порождаемых эмпирических зависимостей на описания новых, ранее не изученных прецедентов.

4. Показана возможность режима параллельных вычислений для реализации метода последовательных приближений.

5. Математически обоснована корректность метода последовательных приближений, для формирования диаграмм вложимости классов эквивалентности прецедентов.

Вместе с тем, работа не свободна от недостатков.

1. В доказательстве Утверждения 2.1.1. содержится ошибка. Отношение $E_{\Omega, \nu}$ не является отношением эквивалентности, так как не обладает рефлексивностью.

2. В условиях Утверждения 5.2.3 (Теоремы о представлении) не указано, что она справедлива лишь в том случае, когда решетка сходств объектов является дистрибутивной.

3. В работе встречается термин «нумерованное множество», хотя это, судя по определению автора, не что иное, как мультимножество. (справедливости ради следует отметить, что в ряде мест автор использует и этот термин).

4. В диссертации, особенно в её обзорных разделах, нередко используются англоязычные термины. Для многих из них существуют общепринятые русские эквиваленты. Но даже в противном случае, одной из задач научной работы является развитие русского научного подязыка.

Однако, отмеченные недостатки не повлияли на результаты работы. Основные результаты диссертации опубликованы в научных изданиях, рекомендованных ВАК Минобрнауки России. Все основные результаты строго доказаны. Работа изложена ясным языком. Совокупность результатов, полученных автором работы, безусловно является научным достижением в области теоретических основ информатики. Диссертационная работа удовлетворяет требованиям ВАК МОН РФ, а её автор Михаил Иванович Забежайло заслуживает присуждения искомой ученой степени доктора физико-математических наук по специальности 05.13.17.-Теоретические основы информатики.

Доктор физико-математических наук

профессор



Г.С.Осипов

«09» июня 2015 года.

Подпись Осипова Г.С. заверяю

Ученый секретарь ИАС РАН
К.ф.-м.н.

Осипов С.Н.

Осипов Геннадий Семенович
8 (499) 135-4222
gos@isa.ru