

## ОТЗЫВ

официального оппонента на диссертацию Виноградова Дмитрия Вячеславовича «Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток», представленную на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 «Теоретические основы информатики»

Диссертационная работа Д.В. Виноградова посвящена применению методов цепей Маркова и прикладной теории решеток для создания и исследования нового метода машинного обучения, который автор называет ВКФ-методом. **Актуальность** темы диссертации не вызывает сомнений, так как в современной теоретической и прикладной информатике методы машинного обучения образуют фундамент. Однако, имеющиеся сейчас методы не лишены недостатков: поход на основе нейронных сетей не может обеспечить объяснения механизмов принятия решений, а комбинаторные методы обучения являются очень трудозатратными.

Вероятностная парадигма, основанная на цепях Маркова, позволила автору избежать необходимости порождения всех сходств, число которых в худшем случае может оказаться экспоненциально большим. Теория решеток, с другой стороны, позволяет свести базовую операцию сходства к побитовому произведению битовых представлений объектов и их фрагментов.

Диссертация Д.В. Виноградова состоит из введения, четырех глав, заключения и списка литературы.

**Первая глава** определяет необходимые понятия и базовые алгоритмы, опираясь на прикладную теорию решеток. Сначала излагается алгоритм кодирования объектов битовыми строками (с заменой операции сходства на побитовое умножение) и доказываемся его корректность и минимальность полученного представления. Основываясь на этом представлении, автор вводит операции «Замыкай-по-одному». После их анализа предлагается ленивая схема их вычислений, и доказываемся теорема о том, насколько реже будет вызываться вычислительно трудная операция «замыкания», если заменить классический алгоритм его ленивым вариантом.

**Вторая глава** посвящена изложению математических результатов автора о проблеме переобучения (внезапно возникших сходствах). Модель, которая здесь используется, определяется через серии Бернулли на признаках, комбинации которых могут образовывать такие сходства. Доказываемся теорема о невозможности устранить переобучение через минимальное число родителей гипотез. Для техники «запрет контр-примеров» (примеров, не обладающих целевым свойством) доказан асимптотический результат и получена оценка скорости сходимости. Этот результат показывает невозможность избавиться от переобучения и в этом случае. Вторая половина главы 2 посвящена выводу явных формул на производящие функции вероятностей преодоления нужного числа контр-примеров при помощи сходства заданного числа родителей.

**Третья глава** посвящена цепям Маркова и основанным на них вероятностным алгоритмам для поиска сходства. Исследованы немонотонная, монотонная, спаривающая, ленивая спаривающая, остановленная ленивая спаривающая цепи Маркова. На примерах показана неравномерность распределения остановки соответствующих алгоритмов. Спаривающие цепи Маркова обладают хорошими свойствами: останавливаемость с вероятностью единица, расстояние между распределениями остановки между остановленным и неостановленным вариантами в метрике тотальной вариации экспоненциально убывает. Для частного случая Булеана удалось получить оценку (и доказать теорему о сильной концентрации) на время нахождения случайного подмножества спаривающей цепью Маркова, отличающейся от прямого задания его последовательностью испытаний Бернулли на логарифмический множитель. Для монотонной цепи Маркова в случае Булеана доказаны теоремы об оценке времени перемешивания и точности этой оценки.

**Четвертая глава** посвящена изложению процедур машинного обучения. После их изложения доказана ключевая теорема об оценке числа гипотез, необходимых для надежного предсказания всех важных положительных примеров. Потом описывается программная реализация и применение программной системы к двум массивам данных. Предложенный автором подход продемонстрировал преимущества над некоторыми алгоритмами символьного обучения и возможность его применения к массивам данных, содержащим несколько тысяч обучающих примеров.

Вероятностный подход к анализу данных на основе операции сходства ранее не исследовался. Исследование феномена переобучения в этой парадигме также является **новым**.

Научные результаты и выводы, сформулированные в диссертации, являются строго **обоснованными**.

Автореферат адекватно отражает содержание диссертационной работы.

#### **Недостатки работы.**

1. Формулировки на стр. 31-32 нуждаются в уточнении. То, что там называется производящей функцией моментов целочисленной случайной величины является просто производящей функцией, так как события, вероятности которых дают коэффициенты, не образуют полную группу событий (опущено значение 0). При этом результат теоремы 1.2 остается верным, так как вероятность нулевого значения не дает вклада в среднее.
2. При описании на странице 110 диссертации Алгоритма 8 (ленивый вариант спаривающей цепи Маркова) опущены операторы присваивания новых значений для переменной *moveUp* (внутри первого if должно быть *moveUp := false*, а внутри третьего – *moveUp := true*).

Перечисленные недостатки не снижают ценности диссертационной работы Виноградова Д.В. и не влияют на ее положительную оценку.

Диссертационная работа Д.В. Виноградова «Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток» **соответствует критериям**, установленным «Положением о присуждении ученых степеней», утвержденным Постановлением Правительства РФ от 24 сентября 2013 г. № 842.

Автором исследований разработаны теоретические положения, совокупность которых можно квалифицировать как **открытие нового научного направления** «машинное обучение, основанное на теории решеток» в теоретической информатике.

Данная работа **полностью соответствует** всем требованиям ВАК РФ, предъявляемым к диссертациям на соискание ученой степени доктора физико-математических наук, а ее автор – Виноградов Дмитрий Вячеславович – **заслуживает** присуждения ему ученой степени доктора физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики».

Официальный оппонент:

д.ф.-м.н., доцент,

дата: 07.02.2019

М.Н. Устинин

Почтовый адрес:

125047, Москва, Миусская пл., д.4, ИПМ им. М.В. Келдыша РАН

Телефон: +7(4967)31-85-04; +7(499)978-13-14

Организация – место работы: Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»

Должность: Зам. директора по филиалу

Адрес электронной почты: [ustinin@impb.ru](mailto:ustinin@impb.ru)

Веб-сайт: <http://www.keldysh.ru/>

Подпись М.Н.Устинина заверяю:

Ученый секретарь Федерального государственного учреждения «Федеральный исследовательский центр Институт прикладной математики им. М.В.Келдыша Российской академии наук»

к.ф.-м.н.

А.И. Маслов

дата:

07.02.2019

