

ОТЗЫВ

официального оппонента на диссертацию Стрижова Вадима Викторовича «Порождение и выбор моделей в задачах регрессии и классификации», представленную на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 – теоретические основы информатики

Решается задача построения и обоснования методов выбора моделей из индуктивно порождаемого множества. Идея предлагаемого подхода заключается в снижении мощности множества, из которого производится выбор моделей. Рассматриваются методы последовательного порождения моделей оптимальной сложности. Для выбора и оценки сложности моделей используется связанный байесовский вывод. Исследуются функция правдоподобия выборки, априорные и апостериорные распределения параметров и функция правдоподобия модели. Предлагается ряд способов получения оценок параметров и гиперпараметров моделей. Полученные оценки используются в процедуре направленного поиска оптимальной структуры модели. Они указывают, какие элементы модели являются наиболее информативными. Одним из требований, накладываемых на модель, является ее интерпретируемость: решается задача построения такой модели, структура которой была бы содержательно оценена экспертами в предметной области. Предлагаемые методы построения моделей регрессии, классификации и прогнозирования решают широкий класс задач машинного обучения.

Задача выбора моделей, решаемая в диссертационной работе, является одной из центральных и актуальных проблем машинного обучения и интеллектуального анализа данных. В работе предложен новый подход к автоматическому построению моделей регрессии и классификации. Центральной проблемой, исследуемой в работе, является проблема выбора из большого количества порождаемых моделей. Для решения этой проблемы разработаны новые методы направленного поиска в пространстве моделей. Приводится теоретическое обоснование предлагаемых методов. Достоверность методов подтверждается результатами вычислительных экспериментов, приведенных в работе. Таким образом, предлагаемый подход является новым, а форма представления материала работы и полученные выводы – достоверными. Научные положения, выводы и рекомендации, сформулированные в диссертации, являются обоснованными.

Первая глава посвящена постановке задачи математического моделирования измеряемых данных и основным вопросам, возникающим при исследовании свойств. Предлагается рассмотреть задачу математического моделирования как задачу восстановления регрессии. Сформулированы основные задачи регрессионного анализа, связанные с выбором модели и оценкой ее параметров; перечислены методы их решения.

Вторая глава посвящена методам порождения параметрических моделей. Перечислены алгоритмы порождения моделей заданной сложности с указанием

недостатков, на устранение которых направлены предлагаемые в работе методы. Описаны условия допустимости порождаемых моделей, способы модификации и упрощения суперпозиции путем замены на изоморфную ей суперпозицию меньшей сложности. Предлагается метод прогнозирования структуры модели по набору прецедентов – пар выборка-модель, позволяющий значительно упростить процедуру порождения моделей.

В третьей главе рассмотрены подходы к выбору модели из порождаемого множества: выбор из множества всех допустимых суперпозиций ограниченной сложности или последовательный выбор моделей при их порождении. В работе развивается второй подход. Он частично снимает необходимость решения некорректно поставленной задачи при оптимизации параметров моделей. При таком подходе предполагается, что модель представима в виде композиции ее элементов. Тогда последовательным добавлением или удалением элементов достигается оптимальное соотношение между качеством модели и ее структурной сложностью. Приведен анализ алгоритмов выбора признаков в линейной регрессии. Предложен метод связанного отбора признаков и выбора модели. Процедура поиска оптимального решения представлена в виде пути в n -мерном кубе.

В четвертой главе рассмотрен подход к решению задачи выбора моделей, основанный на максимизации правдоподобия – связанный байесовский вывод. Установлена связь между частотным и байесовским подходами к оценке правдоподобия модели. Для ряда гипотез о распределении данных предложены методы оценки гиперпараметров модели, используемых при выборе признаков и модификации структуры модели. Приведен пример применения предлагаемых методов к отбору шумовых и мультикоррелирующих признаков; к построению многоуровневых моделей и смесей моделей.

В пятой главе рассмотрены способы выбора моделей в задачах построения интегральных индикаторов качества объектов, заданных экспертами в ранговых шкалах. Сформулировано условие согласованности решения: существование непустого множества – пересечения образа весов показателей и допустимого множества интегральных оценок. Предложены итерационный алгоритм построения индикаторов, описана процедура их уточнения в случае, если условие согласованности не выполнено.

В шестой главе рассмотрен ряд прикладных задач, решаемых в рамках задачи регрессионного анализа, в том числе методами, описанными в работе выше.

Работа не свободна от недостатков, не снижающих, впрочем, положительного впечатления о ней: недостаточно подробно изучены свойства оценок вероятностей связей между элементами суперпозиции, не проведен сравнительный анализ предложенных способов получения оценок гиперпараметров моделей, в работе не исследуется проблема многоэкстремальности пространства параметров моделей.

Результаты диссертационной работы описаны в научных статьях, опубликованных в российских и международных журналах, входящих в перечень

ВАК, и представлены на российских и международных конференциях. Диссертационная работа соответствует критериям, установленным Положением о присуждении ученых степеней: решена задача, имеющая значение для развития методов машинного обучения и анализа данных, она написана автором самостоятельно, обладает целостностью и содержит новые научные результаты, полученные автором лично. Приведены рекомендации по использованию научных выводов. Работа полностью соответствует требованиям ВАК РФ, предъявляемым к диссертациям на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 — теоретические основы информатики, а ее автор В. В. Стрижов заслуживает присуждения ему ученой степени доктора физико-математических наук по данной специальности.

Официальный оппонент
д.ф.-м.н.



Ю. Г. Сметанин

*Подпись Сметанина Ю. Г.
завершено.*

Главный специалист по картам



*И. А. Мельникова
23 октября 2014 г.*

Сметанин Юрий Геннадиевич

Доктор физико-математических наук, начальник Отдела инфокоммуникационных технологий и вычислительных систем Российского фонда фундаментальных исследований

Адрес: 119991, Москва, Ленинский проспект, 32а, 20-21 этаж, В-334, ГСП-1

Тел.: 8 495 952 50 25, Email: smetanin@rfbr.ru