

Отзыв официального оппонента

на диссертационную работу Зайцева Алексея Алексеевича
«Методы построения регрессионных моделей разнородных источников данных для
индустриальной инженерии», представленную на соискание ученой степени кандидата
физико-математических наук по специальности 05.13.18 – математическое моделирование,
численные методы и комплексы программ.

В работе Зайцева А.А. рассматривается задача регрессии, которая часто встречается в различных областях прикладной математики. Автор рассматривает специальный случай этой общей задачи: в работе предполагается, что данные поступают из разных источников, отличающихся по стоимости доступа к данным и погрешности их получения. Как правило, источники бывают двух видов: источник, предоставляющий точные и дорогие данные и источник позволяющий получить данные существенно дешевле, но, как правило, с гораздо большей погрешностью. Как правило, дешевизна альтернативного источника данных позволяет получить выборки существенно большего размера, что необходимо учитывать при построении регрессионной. В работе решается задача построения регрессионной зависимости по обучающей выборке для случая, когда обучающая выборка поступает из разнородных источников. Исследование посвящено задаче выбора соотношения размеров выборок, полученных из разнородных источников данных, позволяющих получить модель наилучшего качества в рамках заданного ресурсного ограничения. В качестве метода построения регрессионных моделей используется нелинейная регрессия на основе гауссовских процессов, которая позволяет естественным образом учесть разнородность данных. Актуальность подобной постановки задачи обусловлена широким использованием данных, полученных из разнородных источников, в задачах индустриальной инженерии.

В главе 1 даны примеры прикладных задач индустриальной инженерии, в которых доступны разнородные источники данных, обоснован выбор регрессии на основе гауссовских процессов в качестве исследуемого метода для построения нелинейных регрессионных моделей. Затем дана формальная математическая постановка задачи, в которой в частности зафиксирована линейная модель связи между разнородными источниками данных – такая модель обычно рассматривается как при решении прикладных задач, так и в теоретических исследованиях.

Глава 2 посвящена описания теоретических свойств регрессии на основе гауссовских процессов и проблемах практического использования этого подхода для случая однородных данных. Показано, что и с теоретической, и с практической точки зрения представляется

полезным использовать байесовский подход для оценки параметров модели. При этих предположениях в работе даются теоретические оценки ошибки интерполяции для случая известных спектральных плотностей и случая, когда спектральные плотности процессов неизвестны, но удовлетворяют минимаксным ограничениям. Также проведено тестирование использования байесовского подхода на широком наборе искусственных задач. Показано, что байесовский подход позволяет избежать проблем построения регрессии на основе гауссовских процессов, связанных с переобучением модели.

В главе 3 диссертант предлагает алгоритмы оценивания регрессионных моделей на основе гауссовских процессов для разнородных данных, обладающие низкой вычислительной сложностью. В главе описывают существующие методы оценивания регрессионной модели на основе гауссовских процессов, выясняется их недостаточность для случая больших обучающих совокупностей, а затем предлагается подход, основанный на низкоранговой аппроксимации ковариационных матриц. Производится теоритическая оценка вычислительной сложности и качества предложенных приближенных алгоритмов. Глава также содержит описание комплекса программ, в котором реализованы предложенные методы.

В главе 4 приведены теоретические оценки качества регрессионной модели на основе гауссовских процессов для случая множественной регрессии при известной корреляционной функции для однородных и разнородных данных. В главе также приводится описание метода выбора соотношения между размерами выборок разнородных данных.

Глава 5 посвящена описанию результатов вычислительных экспериментов, проведенных с использованием разработанного комплекса программ. Эксперименты на реальных и модельных данных показывают, что разработанная в диссертационном исследовании методика позволяет получать регрессионные модели высокого качества по сравнению с эвристическими методами, существующими в литературе.

Вместе с тем надо отметить несколько недостатков работы:

1. В разделе 4 рассматриваются выборки данных на бесконечной решетке, в то время как предполагается использовать предложенный метод для выборок конечного размера.
2. В параграфе 5.2. показано, что предложенные в диссертации подходы могут использоваться в случае наличия более чем двух источников данных, однако, недостаточно подробно описано, как происходит в этом случае оценка параметров регрессионной модели.
3. В исследовании даны оценки неопределенности для регрессионной модели, полученной при использовании низкоранговой аппроксимации ковариационной матрицы, однако не приведена экспериментальная проверка качества такой оценки неопределенности на реальных задачах.

Важно заметить, что указанные недостатки не уменьшают ценности работы и ее общей положительной оценки. Диссертация Зайцева А.А. представляет собой законченную научно-квалификационную работу, в которой решена актуальная для прикладных задач инженерного проектирования задача построения аппроксимации по выборкам, полученным из разнородных источников данных, и которая содержит новые научные результаты, обладающие несомненной практической и теоретической ценностью. Достоверность полученных результатов обеспечена как строгими доказательствами, основанными на использовании хорошо изученных методов математической статистики и спектрального анализа, так и результатами вычислительных экспериментов. Автореферат полностью отражает содержание основного текста.

Считаю, что диссертация удовлетворяет требованиям п. 9 «Положения о порядке присуждения ученых степеней» ВАК РФ, а ее автор Зайцев Алексей Алексеевич заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ.

Отзыв подготовил доцент кафедры математических методов прогнозирования факультета ВМК МГУ,
кандидат физ.-мат. наук, доцент
Красоткина Ольга Вячеславовна

Телефоны мобильный: +79156870606

E-mail: o.v.krasotkina@cs.msu.ru

Официальный оппонент, к.ф.-м.н., доцент
кафедры математических методов
прогнозирования факультета ВМК МГУ
имени М.В. Ломоносова

О.В. Красоткина

Ученый секретарь кафедры математических методов прогнозирования факультета ВМК
МГУ имени М.В.Ломоносова,
н.с.

Д.А. Кропотов

Подпись официального оппонента заверяю:
декан факультета ВМК МГУ
имени М.В. Ломоносова,
академик РАН



Е.И. Моисеев