

## **Отзыв**

на автореферат диссертации Зайцева Алексея Алексеевича  
«Методы построения регрессионных моделей разнородных источников данных для  
индустриальной инженерии»,  
представленной на соискание учёной степени кандидата физико-математических наук.  
Специальность 05.13.18 – Математическое моделирование, численные методы и комплексы  
программ.

В последнее время большую популярность приобретает технология предсказательного моделирования, позволяющая строить новые вычислительно эффективные модели по результатам вычислительных экспериментов с исходной моделью. Математическая модель  $M$  проектируемого объекта может быть описана в виде функциональной зависимости  $y = F(x) \in \mathbb{R}^n$ , где переменная  $x \in X \subset \mathbb{R}^p$  описывает сам объект, а так же, если необходимо, может описывать условия его функционирования, параметры управления объектом, параметры внешней среды и т.д. Функция  $F$  является неизвестной, и для ее вычисления проводятся различные натурные и вычислительные эксперименты. Как правило, эксперименты, обеспечивающие приемлемую точность решения, являются трудоемкими, их мало и полученных результатов не хватает для построения адекватных моделей. Однако в любой инженерной области существуют упрощенные модели вычислений, которые позволяют получить менее точные результаты, но существенно быстрее и в достаточно большом количестве. Поэтому в настоящее время большой интерес представляет задача построения аппроксимационной модели  $M'$  путем обработки имеющихся разноточных данных так, чтобы модель  $M'$  приближала исходную математическую модель  $M$  с приемлемой точностью.

В настоящей работе рассматривается линейная модель связи разноточных данных, где целевая функция, аппроксимирующая источник точных данных, есть линейная комбинация функций данных низкой точности и функции поправки. Предполагается, что функции высокой и низкой точности есть реализации независимых стационарных гауссовских процессов, а исходными данными являются наблюдения значений этих функций на некотором наборе точек. Цель — построить регрессионную модель  $M'$  источника данных высокой точности.

Регрессия на основе гауссовских процессов широко используется для построения нелинейных регрессионных моделей в прикладных задачах. Очевидно, что с теоретической точки зрения большой интерес представляет оценка качества регрессионной модели по всей области дизайна. В качестве оценки качества рассматривается минимаксная ошибка — ошибка наилучшей аппроксимации для наихудшей целевой функции. Задача оценки качества модели решалась ранее для однородных данных, однако, наиболее полно был исследован только одномерный случай. В работе получена теоретическая оценка качества регрессионной модели для многомерных нелинейных зависимостей как для случая одного источника данных, так и для множества источников. Этот результат является новым.

Исходными данными задачи построения аппроксимационной модели  $M'$  являются выборки разноточных данных. Одна из таких выборок есть результаты экспериментов высокой точности, другие — низкой. Данных низкой точности может быть много, но все они содержат некоторую погрешность. Следовательно, количество таких данных в обучающей выборке должно быть как-то регламентировано, чтобы эта погрешность в целевой модели не стала значительной. Поэтому следующей актуальной задачей является задача определения плана эксперимента для получения данных низкой точности, а также определения соотношения между объемами выборок разноточных данных. Эта задача является новой и успешно решена автором. В работе предложен и обоснован метод выбора соотношения между размерами выборок разнородных данных. Доказано, что предложенный метод является теоретически оптимальным

и на практике обеспечивает высокое качество регрессионных моделей. Разработан алгоритм выбора оптимального соотношения размеров выборок разнородных данных.

Как уже упоминалось выше, данных низкой точности может быть слишком много, когда построение модели становится невозможным из-за ограничений по вычислительным ресурсам (памяти и времени). Автор предложил эффективный метод построения регрессионной модели на основе низкоранговой аппроксимации.

Таким образом можно сделать вывод, что актуальность и новизна работы соискателя не вызывает никаких сомнений.

В главе 1 автор описывает ряд прикладных задач, мотивирующих решение сформулированных в диссертации задач, приводит формальную математическую модель. Оригинальной является теоретическая формулировка задачи выбора оптимального сеточного плана экспериментов в смысле минимизации минимаксной ошибки интерполяции.

В главе 2 подробнее рассматривается процедура построения модели регрессии на основе гауссовских процессов, проанализирована эффективность байесовского метода оценки как с теоретической, так и с практической точки зрения.

В главе 3 приводится оригинальный алгоритм построения модели по разнородным данным, который является вычислительно более эффективным, чем алгоритмы, известные ранее. Получены теоретические оценки качества этого алгоритма, описан комплекс программ, созданный на основе результатов, полученных в диссертационном исследовании.

В главе 4 приведены основные теоретические результаты работы. Получена минимаксная ошибка интерполяции в многомерном случае как для одного, так и для множества разнородных источников данных. Разработан алгоритм выбора плана эксперимента, минимизирующего минимаксную ошибку интерполяции для разнородных данных.

В главе 5 приведены результаты применения предложенных в исследовании алгоритмов для решения реальных задач.

Обоснованность результатов, выдвинутых соискателем, основывается на согласованности данных эксперимента и научных выводов. Так, в главе 5 автор приводит результаты вычислительных экспериментов, в которых он исследует свойства предложенных алгоритмов и методов на искусственных данных и на данных из реальных инженерных задач.

Таким образом, можно сделать вывод, что в целом, результаты, полученные автором, являются новыми научными знаниями в области математического моделирования.

Принятые в работе допущения и ограничения обоснованы и отражены в полном объеме. Достигнута достаточная точность расчетов. Профессионально применяются методы математического анализа и математической статистики. Работа представляется актуальной, выполнена в полном объеме на достаточном научном уровне. Полученные автором результаты достоверны, выводы и заключения обоснованы.

Исходя из представленных в автореферате сведений, диссертация написана на высоком научном уровне, соответствует требованиям ВАК Минобрнауки России, и соискатель Зайцев Алексей Алексеевич заслуживает присуждения ученой степени кандидата физико-математических наук.

Чернова Светлана Станиславовна  
к.т.н., руководитель отдела тестирования,  
компания DATAADVANCE  
[svetlana.chernova@datadvance.net](mailto:svetlana.chernova@datadvance.net)  
ООО Датадванс, 117246, Москва, Научный проезд, д. 17, 15 этаж. Тел.: +7 (495) 669-68-15,  
[www.datadvance.net](http://www.datadvance.net), [info@datadvance.net](mailto:info@datadvance.net)

Подпись Черновой Светланы Станиславовны заверяю.

 18.05.2017

