

ВОССТАНОВЛЕНИЕ РЕГРЕССИОННЫХ ЗАВИСИМОСТЕЙ В ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ С ПОМОЩЬЮ МЕТОДИКИ СИНТЕЗА АППРОКСИМИРУЮЩЕЙ ФУНКЦИИ

А.Б. Иващенко

Донецкий национальный технический университет

Представлено ефективну методику синтезу апроксимуючої функції. Описано етапи алгоритму: формування банку функцій, відбір перспективних функцій і елімінація непотрібних. Обговорюються особливості програмної реалізації і результати тестування алгоритму. Виявлено недоліки і переваги методики. Висунуто пропозиції з приводу вдосконалення методики.

При обработке и анализе больших объемов информации, представленной в табличном виде, возникает необходимость выявления определенной зависимости между переменными, которую было бы можно представить с помощью математической формулировки. Целью такого поиска может быть как изучение степени влияния одних переменных (факторов) на другие (отклики), так и использование данной зависимости при составлении прогноза поведения исследуемой величины. Естественно, что точность и адекватность выбранной математической модели являются очень важными критериями при выборе ее структуры, то есть вида аппроксимирующих функций, которыми и предполагается описывать зависимость между параметрами [1].

Особенно сложно определится с видом регрессионной модели в случае большого числа независимых переменных. С другой стороны, табличные данные получены обычно экспериментальным путем, поэтому изначально предполагается неустранимая погрешность вычислений. Ввиду этих обстоятельств на практике оказывается очень сложным подобрать одновременно «красивую» и удачно описывающую экспериментальные данные модель.

Существует множество методов аппроксимации, основными среди которых можно считать линейную, полиномиальную и экспоненциальную регрессии, аппроксимацию сплайнами, методы группового учета аргументов, аппроксимация тригонометрическими функциями (спектральный анализ, БПФ, вейвлеты). Описанию и анализу методов регрессионного анализа и построения аппроксимирующих зависимостей посвящено достаточное количество

работ, например [2]. Заметим лишь, что каждый из перечисленных методов «имеет право на существование» и выбор конкретного из них, обычно, обусловлен изначальными предположениями о характере зависимости между величинами. Гораздо сложнее подобрать подходящий по всем критериям метод в случае полного неведения о структуре модели (модели типа «черный ящик»).

Особого внимания заслуживает метод аппроксимации, изложенный в работах В. О. Эглайса [1, 3]. Следует отметить, что с помощью этой методики проведен ряд успешных исследований в отдельных областях науки и техники [4–6].

На основании обзора работ [3–6] был сделан вывод, что применение методики синтеза регрессионных моделей, предложенной В. О. Эглайсом, эффективно и особенно целесообразно, когда структура модели априорно неизвестна.

Целью данной статьи является описание методики синтеза аппроксимирующей функции как полезного и перспективного подхода для восстановления математической модели по табличным данным.

Задача сформулирована следующим образом. Пусть в таблице 1 представлена информация об исследуемой модели.

Таблица 1

Исходные данные

№п/п	x_1	x_2	...	x_n	y
1					
2					
...					
k					

Здесь x_1, \dots, x_n – параметры объекта (или факторы, независимые переменные); y – отклик (зависимая переменная); k – число экспериментов.

Требуется синтезировать соответствующее уравнение регрессии в виде:

$$y = A_0 + F(A_j, X_i) \quad (1)$$

где A_0 – свободный член (постоянная), A_j – набор коэффициентов уравнения регрессии, X_i – набор параметров объекта.

Вкратце, суть методики [1] заключается в следующем. Аппроксимирующая функция строится в классе степенных разложений, и процесс ее нахождения включает следующие этапы. Первоначально, по заданной степени одночлена, формируется

множество базисных функций, затем из этого множества выбирается заданное количество перспективных функций, после чего направленным перебором, который В.Эглайс называет элиминацией, осуществляется окончательный подбор функции адекватной исходной информации.

Описанная методика синтеза регрессионных уравнений была реализована в среде пакета Matlab (рисунок 1). Основной задачей при проектировании программного обеспечения был анализ особенностей программной реализации отдельных этапов и проверка эффективности метода.

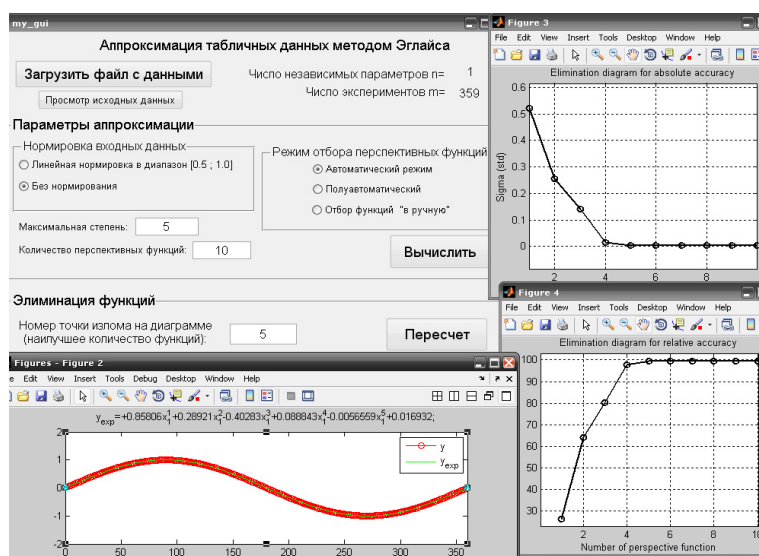


Рис. 1. Интерфейс программного комплекса для синтеза аппроксимирующих функций

Обратим внимание, что с целью исключения деления на ноль в процессе вычислений, перед формированием банка функций была предусмотрена линейная нормировка аргументов:

$$x^* = \frac{(x_{\max}^* - x_{\min}^*)x + (x_{\min}^* x_{\max} - x_{\max}^* x_{\min})}{x_{\max} - x_{\min}} \quad (2)$$

где x – натуральное, то есть исходное, значение независимой переменной, x^* – ее нормированное значение, x_{\min} , x_{\max} – соответственно, минимальное и максимальное значение переменной x ; x_{\min}^* , x_{\max}^* – диапазон изменения ее нормированного значения. Здесь предполагается, что $x_{\min}^* > 0$, $x_{\max}^* > 0$.

Для тестирования и верификации разработанного алгоритма, использовались тестовые наборы «псевдоэкспериментальных» данных с различным числом точек и различным числом входных параметров.

Под этим понималось искусственное задание значений параметров и соответствующих им откликов, аналитический вид зависимости между которыми задан заранее.

Полученные результаты, в целом, подтверждают эффективность рассматриваемого подхода и демонстрируют соответствие заявленным требованиям. Испытание алгоритма и эксплуатация его программной разработки дает возможность говорить как о ряде положительных сторон: универсальности, точности и надежности полученных уравнений, простоте и удобстве пользования методикой, так и о некоторых недостатках. В частности, не во всех случаях алгоритм оправдывает ожидания при восстановлении относительно простых псевдоэкспериментальных зависимостей, в ходе тестирования имели место некоторые сбои, то есть случаи «непопадания» функций, заведомо присутствующих в структуре регрессионной модели, в число перспективных функций. Причины этого, на данный момент, однозначно выяснить не удалось и пока открыты для обсуждения. Наиболее правдоподобной представляется версия, связанная с объемом и количественными особенностями экспериментальных данных.

Кроме этого, тестирование показало, что нормирование исходных данных также, в ряде случаев, может негативно влиять на точность, а, иногда, и на правильность работы программы.

С учетом изложенного представляются целесообразными следующие направления развития данного метода:

- расширение возможности пользователя за счет выбора режима работы алгоритма: с нормировкой или без, или же автоматической проверки входных данных и нормировки их лишь в случае наличия отрицательных;
- обеспечение возможности включения пользователем дополнительных функций из числа базовых в число перспективных, по его усмотрению;
- модификация алгоритма в части совершенствования методики отбора перспективных функций.

Выводы

Рассмотренный метод синтеза аппроксимирующей зависимости экспериментальных данных обладает множеством привлекательных характеристик:

- не требует предварительных знаний о структуре модели;
- данную методику можно считать относительно универсальной, поскольку в качестве базовых функций возможно использование различных по структуре элементарных функций;

– позволяет синтезировать аппроксимирующую функцию из различных комбинаций базовых функций;

– восстанавливаемая модель, в некотором смысле компромиссная, стремится одновременно удовлетворить двум взаимоисключающим оценкам качества модели: критерию точности и показателю эффективности модели, так как позволяет добиться высокой точности и параллельно свести к минимуму число функций-признаков, необходимых для описания отклика и восстановления адекватной регрессионной модели.

Тестирование алгоритма и экспериментальная эксплуатация его программной реализации позволяют говорить о целом ряде положительных сторон рассмотренного подхода: универсальности, точности и надежности полученных уравнений, простоте и удобстве пользования, а также, небольших затратах машинного времени. И, несмотря на ряд выявленных недостатков, можно говорить о высокой степени эффективности данного метода и его достаточном соответствии заявленным требованиям. А закладываемый в алгоритм автоматический выбор структуры аппроксимирующего уравнения является весьма обнадеживающим элементом в части возможности его эффективного использования в задачах типа «черный ящик», в частности, при реконструкции уравнений и моделировании по временным рядам. Поэтому дальнейшее развитие метода, безусловно, представляет интерес.

Библиографический список

1. Эглайс В.О. Аппроксимация табличных данных многомерным уравнением регрессии. Вопросы динамики и прочности: Рига, 1981. Вып. 39. – С. 120 – 125.
2. Дрейпер Н, Смит Г. Прикладной регрессионный анализ. 3-е изд. – М.: Диалектика, 2007. – 912 с.
3. Эглайс В.О. Синтез регрессионной модели объекта на основе табличных данных. // Изв. АН Латв. ССР. Сер. физ. и тех. наук, 1980. – № 4. – С. 109 – 112.
4. Эглайс В.О., Новик Я.А., Дирба Я.А. Синтез регрессионных моделей по результатам численного расчета магнитного поля в электрической машине. // Изв. АН Латв. ССР. Сер. физ. и тех. наук, 1980. – № 4. – С. 113 – 118.
5. Эглайс В.О. Синтез одномерного уравнения регрессии на основе экспериментальных данных. // Изв. АН Латв. ССР. Сер. физ. и тех. наук, 1981. – № 3. – С. 104 – 107.
6. Эглайс В.О., Дирба Я.А., Новик Я.А. Синтез формул кривых намагничивания сталей на ЭВМ по экспериментальным данным. // Изв. АН Латв. ССР. Сер. физ. и тех. наук, 1981. – № 4. – С. 108 – 112.