МЕТОДЫ АНАЛИЗА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ С РАЗЛИЧНЫМИ ВИДАМИ НЕОПРЕДЕЛЕННОСТИ

Шишин В.В.

Национальный исследовательский ядерный университет «МИФИ»

В большинстве случаев полученные в ходе научных экспериментов данные являются неточными, т.е. обладают неопределенностью. Неопределенность в данных обусловлена погрешностями, которые могут быть вызваны различными причинами, такими как ошибки средств измерений, ошибки методов измерений, ошибки округления и др. В общем случае, по характеру проявления составляющие неопределенности можно разделить на случайные и неслучайные. В зависимости от источников неопределенности в экспериментальных данных, а также способа представления исходных данных (точными либо неточными значениями) используемые подходы к их обработке можно разделить на три основные группы − вероятностно-статистический подход, подход на основе теории нечетких множеств и подход на основе интервального анализа. Далее указанные подходы рассматриваются на примере построения по экспериментальным данным линейной зависимости величины отклика Y от влияющих на него факторов Xi:

$$Y = A0+A1X1 + A2X2+\ldots+AnXn \qquad (1)$$

Для решения такой задачи в вероятностно-статистическом подходе традиционно используется аппарат регрессионного анализа [1]. Он основан на предположении о случайном характере ошибок в исходных данных. Чаще всего предполагается, что ошибкам подвержены только измеряемые величины отклика Y, но не факторов Xi. Все величины, полученные в экспериментах, представляются точными числовыми значениями. Для построения зависимости (1) в регрессионном анализе чаще всего применяется метод наименьших квадратов. Основным преимуществом такого подхода является его глубокая теоретическая проработка и реализация во множестве статистических программных пакетов.

При использовании нечеткого регрессионного анализа, реализующего подход на основе теории нечетких множеств, предполагается, что ошибки в данных имеют неслучайный характер, отражающий нечеткость исследуемого явления [2]. При этом так же, как и в классическом регрессионном анализе, ошибкам подвержены только значения отклика Y. Экспериментальные значения факторов Xi представляются точными числовыми значениями, а значения отклика Y могут быть представлены как точными значениями, так и нечеткими числами, с помощью задания функции принадлежности. Для построения модели (1) используются различные способы, например, основанные на методе наименьших квадратов или линейном программировании. Суть подхода состоит в том, чтобы построить такую зависимость, которой удовлетворяют все исходные данные. Для этого коэффициенты A0 и Ai в модели (1) выражаются нечеткими числами.

При использовании интервального анализа неопределенность в данных может иметь как случайную, так и неслучайную составляющие [3]. Ошибкам могут быть подвержены как отклик Y, так и факторы Xi. Исходные экспериментальные величины при этом представляется в виде интервалов их возможных значений. На таких интервалах все значения считаются равновозможными. Интервальные оценки измеряемых величин можно получить, например, используя класс точности прибора либо его цену деления. Для построения модели (1) в интервальном подходе применяются методы линейного программирования либо другие специально разработанные методы, например описанные в [3].

Каждый из указанных подходов имеет свои исходные предпосылки и область применения. Выбор наиболее приемлемого из них должен обосновываться и осуществляться исследователем в каждом конкретном случае, в зависимости от источников неопределенности исходных данных и способов их представления.

Список литературы

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2007. – 912 с.
2. Yun-Hsi O. Chang, Bilal M. Ayyub. Fuzzy regression methods – a comparative assessment // Fuzzy Sets and Systems. − 2001. – Vol. 119 (2). – P. 187-203.
3. Вощинин А.П. Интервальный анализ: развитие и перспективы // Заводская лаборатория. Диагностика материалов. – 2002 − Т.68. №.1 − С. 118-126.