РОБАСТНЫЕ МЕТОДЫ ОБРАБОТКИ ДАННЫХ ХИМИЧЕСКОГО ЭКСПЕРИМЕНТА С ПРИМЕНЕНИЕМ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ

Пантелеймонов А.В., Холин Ю.В.
Харьковский национальный университет им. В.Н. Каразина
panteleimonov@univer.kharkov.ua

Данные химического эксперимента подвержены случайным погрешностям, закон распределения которых, как правило, неизвестен. В то же время, обработка экспериментальных данных обычно построена на предположении о нормальном распределении погрешностей. Для борьбы с последствиями нарушения гипотезы о нормальности нужны робастные методы оценивания.

Данная работа показывает возможность применения теории нечетких множеств к построению робастных процедур обработки данных химического эксперимента.

В теории нечетких множеств результат измерения представляют в виде нечеткого числа x_i с функцией принадлежности $m(x_i) \in [0;1]$, $i \in [1;N]$. Аппроксимирующая экспериментальные данные зависимость F_P – это множество, описываемое уравнением с набором нечетких подгоночных параметров P. Функция принадлежности $m_F(x)$ является мерой истинности высказывания «$x \in F_P$» и представляет собой убывающую функцию расстояния $r(x,F_P)$. Для всего множества результатов функция принадлежности записывается как пересечение функций принадлежности составляющих:

$$m_F(P) = \frac{1}{N}\sum_i m_{F_P}(x_i, y_i)$$

Оптимальным набором параметров для зависимости F_P является такой, для которого функция принадлежности максимальна. Если функция принадлежности i-го измерения представляется в виде эллипса с полуосями u_i и v_i, которые могут быть интерпретированы как погрешности в предикторе и отклике соответственно, т.е. уравнением

$$m_i = 1 - \left(\left(\frac{x - x_i}{u_i}\right) + \left(\frac{y - y_i}{v_i}\right)\right),$$

то возможно максимизацией величины $m_F(P)$ оценить параметры регрессионной модели или проверить устойчивость решения. Алгоритм максимизации функции принадлежности по параметрам регрессии был применен к построению прямой с заранее внесенной выпадающей точкой. Оценки метода наименьших квадратов неудовлетворительны, а оценки метода наименьших модулей практически совпадают с оценками, полученными с применением теории нечетких множеств.

В задаче установления метрологических характеристик визуальных тестовых методик часто наблюдается близость значений критериев согласия для испытываемых теоретических статистических функций. Это не дает возможности однозначно определить вид распределения и его параметры. Процедура нахождения максимума функции принадлежности по параметрам распределения дает единственное решение, которое нечувствительно к выпадающим точкам и свободно от априорной информации о виде распределения погрешностей измерений.

Таким образом, нечеткая логика в решении задач анализа данных позволяет получать робастные оценки.