Для анализа результатов физического эксперимента требуется восстановить функциональную зависимость между измеряемыми величинами. При этом требуется возможность экспертной интерпретации полученной зависимости. Во многих случаях вид функциональной зависимости заранее известен, либо необходимо сделать выбор между несколькими вариантами моделей.

Одним из методов, позволяющих строить структурно сложные интерпретируемые модели, является символьная регрессия [2]. Различные приближения сравниваются согласно некоторому функционалу ошибки, при этом оптимизация параметров модели проводится с помощью, например, с помощью алгоритма Левенберга-Марквардта (АЛМ).

Однако при анализе физического эксперимента важное значение имеют не только сами параметры искомой функциональной зависимости, но и погрешность их определения, обусловленная погрешностями измеряемых в эксперименте величин. Для задачи линейной регрессии соответствующая задача точно решена в частном случае, когда погрешность определения регрессора пренебрежимо мала, а дисперсия зависимой переменной во всех экспериментальных точках одинакова. Для более сложного случая нелинейной регрессии и ситуации, когда необходимо учитывать погрешности как регрессора, так и зависимой переменной (возможно, различные в различных точках), подобная задача, насколько нам известно, не рассматривалась.

В настоящей работе соответствующая задача поставлена и предложен метод ее решения, включающий два этапа.

Сначала для данной выборки среди некоторого индуктивно порождаемого множества моделей F находится модель — параметрическая формула *f*, минимизирующая некоторый функционал качества (например, среднеквадратичную ошибку) [1].

Затем фиксируется структурный вид формулы *f*, и генерируется некоторое «достаточно большое» число реализаций нормально распределенной векторной случайной величины **ξ** (размерностью, соответствующей размерности вектора независимых переменных в обучающей выборке) с заданной дисперсией, которые добавляются к объектам исходной обучающей выборки и носят характер случайного шума. Для каждой из зашумленных случайных выборок снова находятся оптимальные параметры формулы *f*. Далее по найденному множеству значений каждого параметра ωi рассчитываются характеристики соответствующего распределения: среднее значение <ωi> и эмпирическое стандартное отклонение *D(*ωi). Процедура повторяется для различных значений дисперсии зашумляющей случайной величины **ξ** , что позволяет определить зависимость стандартного отклонения параметров от значений дисперсии *D***ξ**. Вводится понятие устойчивости модели *f* по переменной *x*j для параметра ωi как отношение относительной погрешности параметра *D(*ωi)/<ωi> к относительной дисперсии независимой переменной *x*i (являющейся i-й компонентой признакового описания рассматриваемых объектов) *D***ξ** /<*xi*>. Максимум среди всех устойчивостей по всем независимым переменным *x*j для всех параметров ωi называется устойчивостью модели *f*.

В вычислительном эксперименте предложенный метод был применен для восстановления зависимости показателя преломления полимера n от длины волны λ. Оказалось, что для более корректной с физической точки зрения модельной формулы значение устойчивости и скорость ее изменения при увеличении экспериментальных ошибок измерения n и λ оказываются меньше, чем для некорректных моделей.

Введенную характеристику регрессионной модели (устойчивость) возможно использовать как критерий для выбора моделей наряду с такими параметрами, как структурная сложность модели и значение функционала качества.

1. *Рудой Г.И., Стрижов В.В.* Алгоритмы индуктивного порождения суперпозиций для аппроксимации измеряемых данных. — Информатика и ее применения. — 2013. — № 7. — С. 44-53.
2. *Sammut C., Webb G.I.* Symbolic Regression. — Berlin: Springer, 2010. — 954 с.