

## ПОИСК РЕГРЕССИОННЫХ МОДЕЛЕЙ НА МНОЖЕСТВЕ СУПЕРПОЗИЦИЙ ГЛАДКИХ ФУНКЦИЙ

Стрижов В.В.

Вычислительный центр им. А.А. Дородницына РАН,  
Россия, 119991, Москва, ул. Вавилова, 40, тел. 135–4163, e-mail: strijov@ccas.ru

Алгоритм поиска построен на основе метода группового учета аргументов А.Г. Ивахненко [1] и Байесовского подхода в нейронных сетях с анализом весовых коэффициентов, описанного в статьях С.М. Бишопа [2].

Задача нахождения регрессионной модели нескольких свободных переменных поставлена следующим образом. Задана выборка – множество  $\{\mathbf{x}_1, \dots, \mathbf{x}_N \mid \mathbf{x} \in \mathbf{R}^M\}$  значений свободных переменных и множество  $\{y_1, \dots, y_N \mid y \in \mathbf{R}\}$  соответствующих им значений зависимой переменной. Задано множество  $G = \{g \mid g: \mathbf{R} \times \dots \times \mathbf{R} \rightarrow \mathbf{R}\}$  гладких параметрических функций  $g = g(w, ;, ;, \dots)$ . Первый аргумент функции  $g$  – вектор параметров, последующие – элементы вектора свободных переменных. Рассматривается произвольная суперпозиция, состоящая из не более чем  $r$  функций  $g$ . Эта суперпозиция задает параметрическую регрессионную модель  $f = f(\mathbf{w}, \mathbf{x})$ , которая зависит от вектора параметров и от вектора свободных переменных. Требуется найти такую модель  $f_i$ , которая доставляет максимум функционала  $p(\mathbf{w} \mid \mathbf{x}, y, f_i)$ .

Предложенная процедура использует элементы Байесовского вывода для оценки параметров моделей и гиперпараметров, которые указывают на важность элементов моделей. На основе информации о важности элементов итерационно порождаются новые модели. Сложность моделей ограничивается автоматически при сравнении моделей. Для порождения наиболее вероятных моделей используется генетический оптимизационный алгоритм.

Метод выбора моделей из неявно заданного множества и анализ важности элементов этих моделей был протестирован на задаче по аппроксимации кривой, полученной в результате измерения давления в камере внутреннего сгорания дизельного двигателя. Получена модель с удовлетворительной погрешностью аппроксимации.

Сложные регрессионные модели, например, нейронные сети при обработке результатов измерений часто имеют большое число параметров и получаются переобученными. Предложенный метод позволяет находить несложные и достаточно точные модели. Работа выполнена при поддержке РФФИ, грант 04-01-00401.

### Литература

1. Malada, H.R., Ivakhnenko, A. G. Inductive Learning Algorithms for Complex Systems Modeling. CRC Press, 1994.
2. Bishop, C.M., Tipping, M.E. Bayesian regression and classification. IOS Press, NATO Science Series III: Computer and Systems Sciences, 2000. Volume 190. P 267–285.