

ВЫБОР ОПТИМАЛЬНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Стрижов В.В., Шакин В.В.

Вычислительный центр им. А. А. Дородницына РАН,
Сектор математического моделирования в экологии и медицине,
Россия, 119991, ГСП-1, Москва, ул. Вавилова 40,
Тел.: (095)135-4163, e-mail: strijov@ccas.ru

Широкое практическое применение метода нелинейной оптимизации Левенберга-Марквардта в регрессионном анализе создало базу для разработки алгоритмов синтеза регрессионных моделей, полученных путем композиции гладких функций. Алгоритм, описанный ниже, основан на методе группового учета аргументов А. Г. Ивахненко.

Пусть заданы обучающая выборка – множество $X = \{x_1, \dots, x_l \mid x \in \mathbf{R}^n\}$, значений свободных переменных, множество $Y = \{y_1, \dots, y_l \mid y \in \mathbf{R}^1\}$, соответствующих им значений зависимой переменной и тестовая выборка $X^* = \{x_1^*, \dots, x_k^*\}$, $Y^* = \{y_1^*, \dots, y_k^*\}$. Объединения множеств $X \cup X^*$, $Y \cup Y^*$ мы рассматриваем как результат измерений при проведении эксперимента, модель которого неизвестна. На множестве зависимых переменных задана метрика $\rho(y, y^*)$, которую мы полагаем показателем качества приближения, доставляемого искомой регрессионной моделью f .

Также задано множество $G = \{g \mid g : \mathbf{R} \times \dots \times \mathbf{R} \rightarrow \mathbf{R}\}$ гладких функций. Первый аргумент функции $g = g(\beta, \cdot)$ – вектор параметров β , последующие – переменные из множества действительных чисел. Требуется найти такую композицию $f = f(b, x) = g_1 \circ \dots \circ g_p$, b – присоединенные вектора β_1, \dots, β_p , которая бы обеспечивала $\min(\rho(y^*, f(b, x^*)))$, при условии $b = \arg \min \rho(y, f(b, x))$.

Данная задача разбивается на две последовательно решаемые подзадачи. Первая – нахождение такого вектора b параметров нелинейной регрессионной модели, который бы обеспечивал наилучшее приближение функции f обучающей выборкой. Для ее решения используются методы нелинейной оптимизации. Вторая задача – нахождение такой функции f , которая бы, имея аргументом вектор b , полученный при решении первой подзадачи, давала бы наилучшее приближение функции тестовой выборкой. Эта задача решается с использованием генетических оптимизационных алгоритмов.

Работа выполнена при поддержке гранта РФФИ 04-01-00401-а.

Литература.

1. Nikolaev, N. and Iba, H. Accelerated Genetic Programming of Polynomials, Genetic Programming and Evolvable Machines, Kluwer Academic Publ., vol.2, N:3, pp.231–257.
2. Стрижов В. В. Шакин В. В. Прогноз и управление в авторегрессионных моделях. Доклады конференции ММРО-11. РАН, ВЦ, РФФИ, Москва, 2003. – С. 178–181.