Journal of Computational Technologies 2007 (1), P. 93–102

The search of a parametric regression model in an inductive-generated set

Strijov, V.
Computing Center of the Russian Academy of Sciences, Moscow, Russia
e-mail: strijov@ccas.ru

Abstract

The procedure of the search for a parametric regression model in a model set is described. The model set is a set of superpositions of given smooth functions. The model parameters density estimates are used for the search. To illustrate the approach, the problem of a pressure in a spray chamber of a combustion engine is included.

ПОИСК ПАРАМЕТРИЧЕСКОЙ РЕГРЕССИОННОЙ МОДЕЛИ В ИНДУКТИВНО ЗАДАННОМ МНОЖЕСТВЕ*

В. В. Стрижов
Вычислительный центр имени А. А. Дородницына РАН, Москва, Россия
e-mail: strijov@ccas.ru

A procedure of the search for a parametric regression model in a model set is described. The model set is a set of superpositions of the given smooth functions. The models' parameters density estimates are used for the search. To illustrate applicability of the approach the problem of the pressure variation in a spray chamber of the combustion engine is examined.

Введение

Проблема отыскания оптимальной параметрической регрессионной модели имеет большую историю, однако продолжает оставаться одной их самых актуальных в области распознавания образов. А.Г. Ивахненко, еще в 1968 году, предложил метод группового учета аргументов [1]. Согласно этому методу модель, доставляющая наилучшее приближение, отыскивается во множестве последовательно порождаемых моделей. В частности, для построения моделей как суперпозиций функций использовались полиномиальные функции, ряды Фурье и некоторые другие функции. А.Г. Ивахненко и его ученики создали ряд алгоритмов синтеза моделей и предложили методы оценки качества моделей.

При порождении конкурирующих моделей появляется задача определения значимости элементов модели. В работе К. Бишопа [2] предложен метод анализа распределения параметров однослойных нейронных сетей посредством гиперпараметров, т. е. параметров аппроксимирующих функций. Для каждого элемента сети оценивается плотность гауссовского распределения его параметров и делается вывод о том, насколько информативен данный элемент исследуемой регрессионной модели.

Ле Кюн предложил метод для модификации моделей, называемый методом оптимального отсечения (optimal brain damage) [3]. Этот метод состоит в исключении некоторых, наименее информативных, элементов регрессионной модели с тем условием, что при этом качество аппроксимации уменьшается незначительно. При исключении отдельных элементов модели становится возможным оценить их вклад по значениям заданной функции качества аппроксимации.

Проблема сравнения и выбора регрессионных моделей получила новое развитие после ряда публикаций Д. МакКая [4–6], предложившего при выборе модели из заданного мно-

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 04-01-00401).

жества использовать не информационные критерии, например AIC — Akakie Information Criterion, а двухуровневый байесовский вывод и правило Оккама. На первом уровне вывода вычисляются плотности вероятностей распределения параметров каждой модели из заданного множества, на втором — правдоподобие моделей. Правило Оккама состоит в том, что вероятность выбора более сложной модели меньше, чем более простой при сравнимом значении функции качества аппроксимации.

Метод, предлагаемый в данной работе, заключается в следующем. Поиск моделей выполняется по итерационной схеме "порождение-выбор" в соответствии с определенными правилами порождения моделей и критерием их выбора. Последовательно порождаются наборы конкурирующих моделей. Каждая модель в наборе является суперпозицией элементов заданного множества гладких параметрических функций. После построения модели каждому элементу суперпозиции ставится в соответствие гиперпараметр. Параметры и гиперпараметры модели последовательно настраиваются. Из набора выбираются наилучшие модели для последующей модификации. При модификации моделей по значениям гиперпараметров делаются выводы о целесообразности включения того или иного элемента в модель следующего порождаемого набора.

Поставим задачу нахождения регрессионной модели нескольких свободных переменных следующим образом. Задана выборка — множество $\{\mathbf{x}_1, ..., \mathbf{x}_N | \mathbf{x} \in \mathbb{R}^M\}$ значений свободных переменных и множество $\{y_1, ..., y_N | y \in \mathbb{R}\}$ соответствующих им значений зависимой переменной. Обозначим оба эти множества как множество исходных данных D. Также задано множество $G = \{g | g : \mathbb{R} \times ... \times \mathbb{R} \longrightarrow \mathbb{R}\}$ гладких параметрических функций $g = g(\mathbf{b}, \cdot, \cdot, ..., \cdot)$. Первый аргумент функции g — вектор-строка параметров $\mathbf{b}$, последующие — переменные из множества действительных чисел, рассматриваемые как элементы вектора свободных переменных. Рассмотрим произвольную суперпозицию, состоящую из не более чем r функций g. Эта суперпозиция задает параметрическую регрессионную модель $f = f(\mathbf{w}, \mathbf{x})$. Модель f зависит от вектора свободных переменных $\mathbf{x}$ и вектора параметров $\mathbf{w}$. Вектор $\mathbf{w} \in \mathbb{R}^W$ состоит из присоединенных вектор-параметров функций $g_1, ..., g_r$, т. е., $\mathbf{w} = \mathbf{b}_1 \vdots \mathbf{b}_2 \vdots ... \vdots \mathbf{b}_r$, где $\vdots$ — знак присоединения векторов. Обозначим через $\Phi = \{f_i\}$ множество всех суперпозиций, индуктивно порожденное элементами множества G.

Требуется найти модель f_i, которая доставляет максимум функционала $p(\mathbf{w}|D, \alpha, \beta, f_i)$. Этот функционал, определяемый далее, включает искомую модель $f_i(\mathbf{w}, \mathbf{x})$ и ее дополнительные параметры α и β.

1. Выбор регрессионных моделей и гипотез порождения данных

Общий подход к сравнению нелинейных моделей заключается в следующем. Рассмотрим набор конкурирующих моделей $f_1, ..., f_M$. Априорная вероятность модели f_i определена как $P(f_i)$. При появлении данных D апостериорная вероятность модели $P(f_i|D)$ может быть найдена по теореме Байеса

$$P(f_i|D) = \frac{P(f_i)p(D|f_i)}{\sum\limits_{i=1}^{M} p(D|f_i)P(f_i)},$$

где $p(D|f_i)$ — функция соответствия модели данным. Знаменатель дроби обеспечивает выполнение условия $\sum\limits_{i=1}^{M} P(f_i|D) = 1$.

Вероятности моделей f_1 и f_2, параметры которых идентифицированы по данным D, сравнимы как

$$\frac{P(f_1|D)}{P(f_2|D)} = \frac{P(f_1)p(D|f_1)}{P(f_2)p(D|f_2)}. \tag{1}$$

Отношение $p(D|f_1)/p(D|f_2)$ есть отношение правдоподобия моделей. Отношение $P(f_1)/P(f_2)$ является априорной оценкой предпочтения одной модели другой. При моделировании отдается предпочтение наиболее простым и устойчивым моделям. Если априорные оценки $P(f_i)$ моделей одинаковы, т. е. нет причины предпочитать одну модель другой, то их необходимо сравнивать по значениям $p(D|f_i)$:

$$p(D|f_i) = \int p(D|\mathbf{w}, f_i)p(\mathbf{w}|f_i)d\mathbf{w}.$$

Апостериорная плотность распределения параметров $\mathbf{w}$ функции f_i при заданной выборке D равна

$$p(\mathbf{w}|D, f_i) = \frac{p(D|\mathbf{w}, f_i)p(\mathbf{w}|f_i)}{p(D|f_i)}, \tag{2}$$

где $p(\mathbf{w}|f_i)$ — априорно заданная плотность вероятности параметров начального приближения; $p(D|\mathbf{w}, f_i)$ — функция правдоподобия параметров модели, а знаменатель $p(D|f_i)$ обеспечивает выполнение условия $\int p(\mathbf{w}|D, f_i)d\mathbf{w} = 1$. Он задан интегралом в пространстве параметров $\int p(\mathbf{w}'|D, f_i)p(\mathbf{w}'|f_i)d\mathbf{w}'$. Формулы (2) и (1) называются формулами байесовского вывода первого и второго уровня.

Рассмотрим регрессию $y = f_i(\mathbf{b}, \mathbf{x}) + \nu$ с аддитивным гауссовским шумом с дисперсией σ_ν и нулевым матожиданием. Тогда плотность вероятности появления данных определяется как

$$p(y|x, \mathbf{w}, \beta, f_i) \triangleq p(D|\mathbf{w}, \beta, f) = \frac{\exp(-\beta E_D(D|\mathbf{w}, f_i))}{Z_D(\beta)},$$

где $\beta = 1/\sigma_\nu^2$. Нормирующий множитель $Z_D(\beta)$ задан выражением

$$Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}}, \tag{3}$$

а взвешенный функционал ошибки в пространстве данных

$$\beta E_D = \frac{\beta}{2}\sum_{n=1}^{N}(f_i(\mathbf{x}_n) - y_n)^2. \tag{4}$$

Введем регуляризующий параметр α, который отвечает за то, насколько хорошо модель должна соответствовать зашумленным данным. Функция плотности вероятности параметров с заданным гиперпараметром α имеет вид

$$p(\mathbf{w}|\alpha, f_i) = \frac{\exp(-\alpha E_W(\mathbf{w}|f_i))}{Z_W(\alpha)},$$

где α — обратная дисперсия распределения параметров, $\alpha = \sigma_{\mathbf{w}}^{-2}$, а нормирующая константа Z_W зависит от дисперсии распределения параметров как

$$Z_W(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{W}{2}}. \tag{5}$$

Требование к малым значениям параметров [7] предполагает гауссовское априорное распределение с нулевым средним:

$$p(\mathbf{w}) = \frac{1}{Z_W}\exp\left(-\frac{\alpha}{2}||\mathbf{w}||^2\right).$$

Так как переменные α и β являются параметрами распределения параметров модели, в дальнейшем будем называть их гиперпараметрами. Исключая нормирующую константу Z_W, которая не зависит от параметров $\mathbf{w}$, и логарифмируя, получаем

$$\alpha E_W = \frac{\alpha}{2}||\mathbf{w}||^2. \tag{6}$$

Эта ошибка регуляризирует параметры, начисляя штраф за их чрезмерно большие значения.

При заданных значениях гиперпараметров α и β выражение (2) для фиксированной функции f_i будет иметь вид

$$p(\mathbf{w}|D,\alpha,\beta) = \frac{p(D|\mathbf{w},\beta)p(\mathbf{w}|\alpha)}{p(D|\alpha,\beta)}.$$

Записывая функцию ошибки в виде $S(\mathbf{w}) = \alpha E_W + \beta E_D$, получаем

$$p(\mathbf{w}|D,\alpha,\beta,f_i) = \frac{\exp(-S(\mathbf{w}|f_i))}{Z_S(\alpha,\beta)}, \tag{7}$$

где Z_S — нормирующий множитель.

2. Нахождение параметров модели

Рассмотрим итеративный алгоритм для определения оптимальных параметров $\mathbf{w}$ и гиперпараметров α, β при заданной модели f_i. Корректный подход заключается в интегрировании всех неизвестных параметров и гиперпараметров. Апостериорное распределение параметров определяется как

$$p(\mathbf{w}|D) = \iint p(\mathbf{w},\alpha,\beta|D)d\alpha d\beta = \iint p(\mathbf{w}|\alpha,\beta,D)p(\alpha,\beta|D)d\alpha d\beta, \tag{8}$$

что требует интегрирования апостериорного распределения параметров $p(\mathbf{w}|\alpha,\beta,D)$ по пространству, размерность которого равна количеству параметров. Вычислительная сложность этого интегрирования весьма велика. Интеграл может быть упрощен при подходящем выборе начальных значений гиперпараметров.

Приближение интеграла заключается в том, что апостериорная плотность распределения гиперпараметров $p(\alpha, \beta|D)$ имеет выраженный пик в окрестности наиболее правдоподобных значений гиперпараметров $\alpha^{\text{MP}}, \beta^{\text{MP}}$. Это приближение известно как аппроксимация Лапласа [8]. При таком допущении интеграл (8) упрощается до

$$p(\mathbf{w}|D) \approx p(\mathbf{w}|\alpha^{\text{MP}}, \beta^{\text{MP}}, D) \iint p(\alpha, \beta|D) d\alpha d\beta \approx p(\mathbf{w}|\alpha^{\text{MP}}, \beta^{\text{MP}}, D).$$

Необходимо найти значения гиперпараметров, которые оптимизируют апостериорную плотность вероятности параметров, а затем выполнить все остальные расчеты, включающие $p(\mathbf{w}|D)$ при фиксированных значениях гиперпараметров.

Для нахождения функционала $p(\mathbf{w}|\alpha, \beta, D)$, который использует апостериорное распределение параметров, рассмотрим аппроксимацию ошибки $S(\mathbf{w})$ на основе рядов Тейлора второго порядка:

$$S(\mathbf{w}) \approx S(\mathbf{w}^{\text{MP}}) + \frac{1}{2}(\mathbf{w} - \mathbf{w}^{\text{MP}})^T A (\mathbf{w} - \mathbf{w}^{\text{MP}}). \tag{9}$$

В выражении (9) нет слагаемого первого порядка, так как предполагается, что $\mathbf{w}^{\text{MP}}$ определяет локальный минимум функции ошибки, т. е.

$$\frac{\partial S(\mathbf{w}^{\text{MP}})}{\partial w_\xi} = 0$$

для всех значений ξ. Матрица A — это матрица Гессе функции ошибок:

$$A = \nabla^2 S(\mathbf{w}^{\text{MP}}) = \beta \nabla^2 E_D(\mathbf{w}^{\text{MP}}) + \alpha I.$$

Обозначим первое слагаемое правой части через H, тогда $A = H + \alpha I$. Подставив полученное приближенное значение $S(\mathbf{w})$ в (7) и обозначив $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}^{\text{MP}}$, получим

$$p(\mathbf{w}|\alpha, \beta, D) = \frac{1}{\hat{Z}_S} \exp\left(-S(\mathbf{w}^{\text{MP}}) - \frac{1}{2}\Delta\mathbf{w}^T A \Delta\mathbf{w}\right).$$

Оценим нормирующую константу $\hat{Z}_S$, необходимую для аппроксимации кривой Гаусса, как

$$\hat{Z}_S = \exp(-S(\mathbf{w}^{\text{MP}}))(2\pi)^{\frac{W}{2}} (\det A)^{-\frac{1}{2}}. \tag{10}$$

Максимизируем функцию $p(D|\alpha, \beta)$, изменяя значения гиперпараметров α и β. Это можно выполнить, интегрируя функцию плотности вероятности данных по пространству параметров $\mathbf{w}$:

$$p(D|\alpha, \beta) = \int p(D|\mathbf{w}, \alpha, \beta) p(\mathbf{w}|\alpha, \beta) d\mathbf{w} = \int p(D|\mathbf{w}, \alpha, \beta) p(\mathbf{w}|\alpha) d\mathbf{w}, \tag{11}$$

где второй интеграл справедлив по причине того, что распределение параметров не зависит от дисперсии шума в силу гипотезы о гауссовском распределении шума. Для упрощения вычислений мы допускаем, что распределение $p(\alpha, \beta)$ является равномерным.

Используя (4), (6), запишем (11) в виде

$$p(D|\alpha, \beta) = \frac{1}{Z_D(\beta)} \frac{1}{Z_D(\alpha)} \int \exp(-S(\mathbf{w}))) d\mathbf{w}.$$

Из (3), (5), (10) и предыдущего выражения получим

$$\ln p(D|\alpha,\beta) = -\alpha E_W^{\text{MP}} - \beta E_D^{\text{MP}} - \frac{1}{2}\ln|A| + \frac{W}{2}\ln\alpha + \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi). \tag{12}$$

Для того чтобы оптимизировать это выражение относительно α, найдем производную

$$\frac{d}{d\alpha}\ln|A| = \frac{d}{d\alpha}\ln\left(\prod_{j=1}^{W}\lambda_j + \alpha\right) = \frac{d}{d\alpha}\sum_{j=1}^{W}\ln(\lambda_j+\alpha) = \sum_{j=1}^{W}\frac{1}{\lambda_j+\alpha} = \operatorname{tr}(A^{-1}). \tag{13}$$

В этом выражении $\lambda_1, ..., \lambda_W$ — собственные значения матрицы H. Так как функция ошибки на данных не является квадратичной функцией параметров, как при линейной или RBF регрессии, непосредственно оптимизировать величину α невозможно, гессиан H не является константой, а зависит от параметров $\mathbf{w}$. Поскольку мы принимаем $A = H + \alpha I$ для вектора $\mathbf{w}^{\text{MP}}$, который зависит от выбора α, собственные значения H косвенным образом зависят от α. Таким образом, формула (13) игнорирует параметры модели.

С использованием этого приближения, производная (12) с учетом α равна

$$\ln p(D|\alpha,\beta) = -E_W^{\text{MP}} - \frac{1}{2}\sum_{j=1}^{W}\frac{1}{\lambda_j+\alpha} + \frac{W}{2\alpha}.$$

Приравнивая последнее выражение к нулю и преобразовывая его, получаем выражение для α:

$$2\alpha E_W^{\text{MP}} = W - \sum_{j=1}^{W}\frac{\alpha}{\lambda_j+\alpha}. \tag{14}$$

Обозначим вычитаемое правой части через γ:

$$\gamma = \sum_{j=1}^{W}\frac{\alpha}{\lambda_j+\alpha}.$$

Те компоненты суммы, в которых $\lambda_j \gg \alpha$, привносят вклад, близкий к единице, а те компоненты суммы, в которых $0 < \lambda_j \ll \alpha$, привносят вклад, близкий к нулю. Таким образом γ может быть интерпретирована как мера числа хорошо обусловленных параметров модели.

Для нахождения гиперпараметра β рассмотрим задачу оптимизации (12). Обозначим через μ_j собственное значение матрицы $\nabla^2 E_D$. Так как $H = \beta\nabla^2 E_D$, то $\lambda_j = \beta\mu_j$, а следовательно,

$$\frac{d\lambda_j}{d\beta} = \mu_j = \frac{\lambda_j}{\beta}.$$

Отсюда

$$\frac{d}{d\beta}\ln|A| = \frac{d}{d\beta}\sum_{j=1}^{W}\ln(\lambda_j+\alpha) = \frac{1}{\beta}\sum_{j=1}^{W}\frac{\lambda_j}{\lambda_j+\alpha}.$$

Дифференцируя, как и в случае нахождения α, мы находим, что оптимальное значение β определено как

$$2\beta E_D^{\text{MP}} = N - \sum_{j=1}^{W}\frac{\lambda_j}{\lambda_j+\alpha} = N - \gamma. \tag{15}$$

Способ вычисления оптимальных значений гиперпараметров α и β описан в следующем разделе.

3. Процедура поиска оптимальной модели

Поиск оптимальной модели происходит на множестве порождаемых моделей на каждой итерации алгоритма. Перед работой алгоритма заданы множество измеряемых данных D и множество гладких функций G. Задан начальный набор конкурирующих моделей $F_0 = \{f_1, ..., f_M | f \in \Phi\}$, в котором каждая модель f_i есть суперпозиция функций $\{g_{ij}\}_{j=1}^{r_i}$. Каждой функции g_{ij} — элементу модели f_i ставится в соответствие гиперпараметр α_{ij}, характеризующий начальную плотность распределения вектора параметров $\mathbf{b}_{ij}$ этой функции. Каждой модели f_i поставлен в соответствие гиперпараметр β_i начального приближения. Параметры начального приближения для i-й модели назначаются исходя из априорного распределения данных, определяемых значением β_i. Далее выполняется последовательность шагов, приведенных ниже, которые повторяются заданное количество раз.

1. Методом сопряженных градиентов [9] минимизируются штрафные функции $S_i(\mathbf{w})$ для каждой модели $f_i, i = 1, ..., M$. Отыскиваются параметры моделей $\mathbf{w}_i^{\text{MP}}$.

2. После нахождения параметров $\mathbf{w}_i^{\text{MP}}$ исходя из (14) и (15) определяются новые значения гиперпараметров — α_{ij}^{new} и β_i^{new}. Гиперпараметр β_i функции f_i вычисляется для всего набора данных и равен

$$\beta_i^{\text{new}} = \frac{N - \gamma_i}{E_D(f_i)}.$$

Гиперпараметр α_{ij} вычисляется для каждой функции g_{ij} из суперпозиции f_i и равен

$$\alpha_{ij}^{\text{new}} = \frac{W - \gamma_i}{E_W(\mathbf{b}_{ij})}.$$

Здесь значения функционалов γ_i и $E_W(\mathbf{b}_{ij})$ вычисляются только для подмножества тех параметров $\mathbf{b}_{ij}$ из множества $\mathbf{w}_i$, которые являются параметрами функции g_{ij}. Изменение гиперпараметров повторяется итерационно до тех пор, пока локальный минимум $S_i(\mathbf{w})$ не останется постоянным.

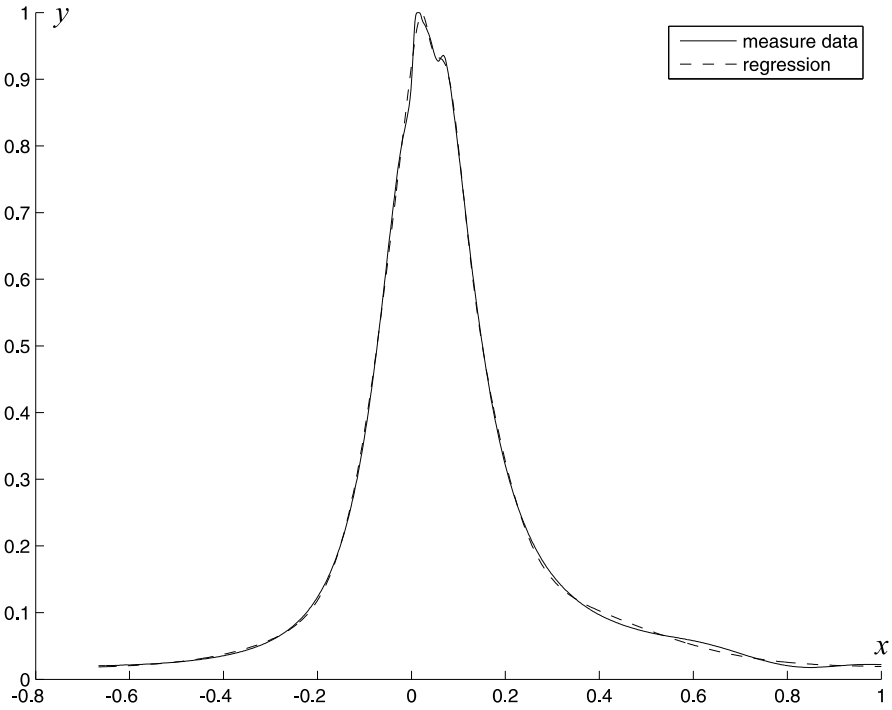
3. Заданы следующие правила построения производных моделей $f_1', ..., f_M'$. Для каждой модели f_i строится производная модель f_i'. В f_i выбирается функция g_{ij} с наименьшим значением α_{ij}. Выбираются произвольная модель f_ξ из $F_0 \backslash \{f_i\}$ и ее произвольная функция $g_{\xi\zeta}$. Модель f' порождается из модели f путем замещения функции g_{ij} с ее аргументами на функцию $g_{\xi\zeta}$ с ее аргументами.

4. С заданной вероятностью η каждая модель f_i' подвергается изменениям. В изменяемой модели выбирается j-я функция, причем закон распределения вероятности выбора функции $p(j)$ задан. Из множества G случайным образом выбирается функция g', которая замещает функцию g_j. Гиперпараметр α_{ij} этой функции определяется как $\max\limits_j(\alpha_{ij})$. Вектор параметров этой функции $\mathbf{b}_{ij}$ равен нулю или назначается при задании G.

5. При выборе моделей из объединенного множества родительских и порожденных моделей в соответствии с критерием $S(\mathbf{w})$ выбираются M наилучших, которые используются в дальнейших итерациях.

4. Численный эксперимент

Ниже описывается пример построения регрессионной модели. Объектом моделирования является кривая одной свободной переменной, представленная набором измерений давления в камере внутреннего сгорания дизельного двигателя. На рисунке сплошной кривой



Исходная выборка и восстановленная выборка, полученная по модели 2.

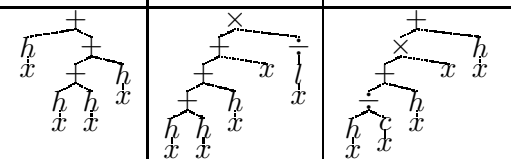
показаны исходные данные, штриховой — значения модели 2. По оси абсцисс отложено значение свободной переменной, по оси ординат — значение зависимой переменной. Выборка, представленная данной кривой, содержит четыре тысячи отсчетов. Для верификации полученных моделей использовалось 118 выборок.

Экспертами задано множество базовых функций G, из элементов которого порождаются регрессионные модели. Список функций приведен в табл. 1. Множество F_0 моделей начального приближения также было задано экспертами.

Т а б л и ц а 1. Множество G базовых функций

№	Функция	Описание	Параметр
Функции двух переменных аргументов, $g(\mathbf{b}, x_1, x_2)$			
1	plus	$y = x_1 + x_2$	—
2	times	$y = x_1 x_2$	—
3	divide	$y = x_1 / x_2$	—
Функции одного переменного аргумента, $g(\mathbf{b}, x_1)$			
4	multiply	$y = ax$	a
5	add	$y = x + a$	a
6	gaussian	$y = \frac{\lambda}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\xi)^2}{2\sigma^2}\right) + a$	λ, σ, ξ, a
7	linear	$y = ax + b$	a, b
8	parabolic	$y = ax^2 + bx + c$	a, b, c
9	cubic	$y = ax^3 + bx^2 + cx + d$	a, b, c, d
10	logsig	$y = \frac{\lambda}{1 + \exp(-\sigma(x-\xi))} + a$	λ, σ, ξ, a

Т а б л и ц а 2. Описание выбранных моделей

Описание	Модель		
	1	2	3
Ошибка ρ_1	0.0034	0.0037	0.0035
Ошибка ρ_2	0.0421	0.0325	0.00338
Число параметров	16	16	16
Структура модели			

Примечание: h — gaussian, c — cubic, l — linear, + — plus, $\times$ — times, $\div$ — divide.

Выбор моделей производился из более тысячи порожденных моделей. В табл. 2 приведены три модели, полученные в результате работы алгоритма. Качество моделей оценивалось по ошибкам ρ_1, ρ_2 и числу параметров в векторе параметров $\mathbf{w}$. Значения ошибок каждой модели получены путем усреднения результатов оптимально настроенной модели по 118 выборкам. Ошибка ρ_1 — среднеквадратичная относительная ошибка

$$\rho_1 = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(\frac{y_i - f(\mathbf{x}_i)}{\max(y_i)}\right)^2},$$

ошибка ρ_2 — максимальная относительная ошибка

$$\rho_2 = \max_{i=1,\ldots,N}\frac{|y_i - f(\mathbf{x}_i)|}{\max(y_i)}.$$

В качестве примера рассмотрим модель 2. Она модель состоит из суперпозиции восьми функций $f_2 = g_1(g_2(g_3(g_4(g_5(x), g_6(x)), g_7(x)), x), g_8(x))$. Функции сложения и умножения $g_1 = \times(\varnothing, \cdot, \cdot)$ и $g_2, \ldots, g_4 = +(\varnothing, \cdot, \cdot)$ имеют первым аргументом пустой вектор параметров; $g_5, \ldots, g_7 = h(\mathbf{b}_i, \cdot)$, $i = 1, \ldots, 3$, и $g_8 = l(\mathbf{b}_4, \cdot)$. Функции $h = \frac{\lambda_i}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x-\xi_i)^2}{2\sigma_i^2}\right)$ имеют векторы параметров $\mathbf{b}_i = \langle\lambda_i, \mu_i, \sigma_i, a_i\rangle$, а функция $l = (ax + b)$ имеет вектор параметров $\mathbf{b}_4 = \langle a, b\rangle$.

Модель f_2 можно переписать в виде

$$f(\mathbf{w}, \mathbf{x}) = l(\mathbf{b}_4, x)^{-1} \times \left(x + \sum_{i=1}^{3} h(\mathbf{b}_i, x)\right),$$

где $\mathbf{x} = x$ и $\mathbf{w} = \mathbf{b}_1 \vdots \mathbf{b}_2 \vdots \mathbf{b}_3 \vdots \mathbf{b}_4$. Развернутый вид модели

$$y = (ax + b)^{-1}\left(x + \sum_{i=1}^{3}\frac{\lambda_i}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x-\xi_i)^2}{2\sigma_i^2}\right) + a_i\right).$$

Модель f_2 была использована экспертами для анализа и прогноза концентрации кислорода в выхлопных газах дизельного двигателя.

Заключение

Универсальные регрессионные модели, например нейронные сети или радиальные базисные функции, при обработке результатов измерений часто имеют большое число параметров и получаются переобученными. Для достижения результатов в построении несложных и достаточно точных моделей поставлена задача о выборе регрессионной модели, которая состоит из суперпозиции гладких функций.

Для выбора наилучшей модели из индуктивно заданного множества использован двухуровневый байесовский вывод. В связи со сложностью вычисления значений интегралов вывода предложены процедуры приближения, которые позволяют отыскивать адекватные модели за приемлемое время вычислений.

Предложенная процедура выбора регрессионных моделей использует гиперпараметры, поставленные в соответствие элементам модели. Эти гиперпараметры указывают на важность элементов модели. На основе информации о важности элементов итеративно порождаются новые модели. Сложность моделей ограничивается автоматически при сравнении моделей.

Описанный метод протестирован на задаче по аппроксимации кривой, построенной в результате измерений давления в камере внутреннего сгорания дизельного двигателя. Получена модель с удовлетворительной погрешностью аппроксимации.

Список литературы

[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 // Advances in Learning Theory: Methods, Models and Applications / J. Suykens, G. Horvath et. al. (Eds). IOS Press, NATO Sci. Ser. III: Computer and Systems Sciences, 2000. Vol. 190. P. 267-–285.

[3] LeCun Y., Denker J.S., Solla S.A. Optimal brain damage // Advances in Neural Information Processing Systems / D.S. Touretzky (Ed.). Morgan Kaufmann, San Mateo, CA, 1990. P. 598–605.

[4] MacKay D. Information, Inference, Learning Algorithms. Cambridge: Cambridge Univ. Press, 2003.

[5] MacKay D. Hyperparameters: optimise or integrate out? // Maximum Entropy and Bayesian Methods / G. Heidberger (Ed.). Santa Barbara, Dordrecht: Kluwer, 1993.

[6] MacKay D. Bayesian interpolation // Neural Comp. 1992. Vol. 4, N 3. P. 415–447.

[7] Nabney I.T. NETLAB: Algorithms for Pattern Recognition. N.Y.; Berlin: Springer-Verl., 2004. P. 330.

[8] MacKay D. Choice of basis for Laplace approximation // Machine Learning. 1998. Vol. 33, N 1.

[9] Branch M.A., Coleman T.F., Li Y. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems // SIAM J. on Sci. Comp. 1999. Vol. 21, N 1. P. 1–23.

Поступила в редакцию 27 июля 2006 г.