

## Регрессия на основе разреженных гауссовских процессов в задаче консолидации разноточных данных\*

*Зайцев А. А., Бурнаев Е. В., Янович Ю. А.*

alexei.zaytsev, evgeny.burnaev, yurij.yanovich@datadvance.net

<sup>1</sup>Москва, ИППИ; <sup>2</sup>Долгопрудный, МФТИ, Premolab; <sup>3</sup>Москва, Datadvance

В работе предложено обобщение подхода для регрессии на основе разреженных гауссовских процессов для решения задачи консолидации разноточных данных. Тестирование подхода проводится на реальной задаче аппроксимации аэродинамических коэффициентов крыла самолета.

### Sparse gaussian processes regression for the data fusion problem\*

*Zaytsev A. A., Burnaev E. V., Yanovich Yu. A.*

ИТР, Moscow, Russia; Premolab, MIPT, Dolgoprudny; Datadvance, Moscow

Sparse gaussian processes regression for the data fusion problem is considered. Proposed approach is tested on a real problem of airfoil coefficients approximation.

В анализе данных рассматривается задача восстановления регрессии [2]. Часто не удается построить линейную модель приемлемого качества, поэтому используются методы, которые позволяют строить нелинейные регрессионные модели [2]. Пример таких моделей — регрессия на основе гауссовских процессов [13].

В некоторых случаях задана выборка значений не только целевой точной функции, но и некоторого ее приближения — грубой функции [14, 11, 7]. В работе рассматривается подход, позволяющий учесть наличие грубого приближения точной функции, основанный на гауссовских процессах. Рассматриваемая задача консолидации точных данных является обобщением задачи построения аппроксимации для функции с многомерным выходом с использованием гауссовских процессов [1]. В работе рассмотрена модель, описанная в работе [6].

Построение нелинейной регрессионной модели, основанной на гауссовских процессах, требует времени кубического по размеру обучающей выборке [13]. Таким образом, в своем исходном виде подход неприменим для выборок больших размеров. Для того, что бы расширить область применимости метода, используется регрессия на основе разреженных гауссовских процессов [13, 8]. В данной работе предлагается подход, позволяющий применять регрессию на основе разреженных гауссовских процессов для решения задачи консолидации разноточных данных.

#### Постановка задачи

Пусть заданы выборки значений для грубой  $D_l = (X_l, y_l) = \{(\mathbf{x}_i^l, y_l(\mathbf{x}_i^l))\}_{i=1}^{N_l}$  и точной  $D_h = (X_h, y_h) = \{(\mathbf{x}_i^h, y_h(\mathbf{x}_i^h))\}_{i=1}^{N_h}$  функции. При этом,

Работа выполнена при поддержке Лаборатории структурных методов анализа данных в предсказательном моделировании, МФТИ, грант правительства РФ дог. 11.G34.31.0073.

$\mathbf{x}_i^l \in \mathbb{R}^n$ ,  $\mathbf{x}_i^h \in \mathbb{R}^n$ ,  $y_l(\mathbf{x}) \in \mathbb{R}$ ,  $y_h(\mathbf{x}) \in \mathbb{R}$ , грубая функция  $y_l(\mathbf{x})$  является некоторым приближением точной функции  $y_h(\mathbf{x})$ .

Используя выборки значений грубой и точной функции, необходимо построить аппроксимацию точной функции  $\hat{y}_h(\mathbf{x}) \approx y_h(\mathbf{x})$ , такую, что для независимой контрольной выборки значений точной функции  $D^* = (X^*, y^*) = \{(\mathbf{x}_i^*, y_h(\mathbf{x}_i^*))\}_{i=1}^{N^*}$  квадратичная ошибка аппроксимации мала:

$$Q(D^*, \hat{y}_h(\mathbf{x})) = \sum_{i=1}^{N^*} (\hat{y}_h(\mathbf{x}_i^*) - y_h(\mathbf{x}_i^*))^2.$$

#### Регрессия на основе гауссовских процессов

Пусть теперь задана выборка значений  $D = (X, y) = \{(\mathbf{x}_i, y(\mathbf{x}_i))\}_{i=1}^N$  и

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x}),$$

где  $f(\mathbf{x})$  — реализация некоторого гауссовского процесса [13],  $\varepsilon(\mathbf{x})$  — белый шум,  $\varepsilon(\mathbf{x}) \sim \mathcal{N}(0, \sigma^2)$ . Тогда  $y(\mathbf{x})$  тоже гауссовский процесс [13], определяемый своим математическим ожиданием  $m(\mathbf{x}) = \mathbb{E}(y(\mathbf{x}))$  и ковариационной функцией  $\text{cov}(y(\mathbf{x}), y(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') = \mathbb{E}(y(\mathbf{x}) - m(\mathbf{x}))(y(\mathbf{x}') - m(\mathbf{x}'))$  [13]. Тогда оценка апостериорного математического ожидания гауссовского процесса для выборки точек  $X^* = \{\mathbf{x}_i^*\}_{i=1}^{N^*}$  имеет вид:

$$\hat{y}(X^*) = K(X^*, X)K^{-1}\mathbf{y},$$

здесь  $K = K(X, X) = \{k(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^N$ ,  $K(X^*, X) = \{k(\mathbf{x}_i^*, \mathbf{x}_j)\}_{i=1, j=1, N^*}$ . Оценка апостериорной ковариационной матрицы точек новой выборки имеет вид

$$\mathbb{V}[X^*] = K(X^*, X^*) - K(X^*, X)K^{-1}K(X, X^*).$$

Вообще говоря, мы не знаем ковариационной функции и математического ожидания гауссовского процесса. Математическое ожидание положим равным нулю, не теряя общности. Пусть теперь ковариационная функция  $\text{cov}(y(\mathbf{x}), y(\mathbf{x}'))$  априори принадлежит параметрическому семейству вида:

$$k(\mathbf{x}, \mathbf{x}' | a, \sigma^2) = \text{cov}(f(\mathbf{x}), f(\mathbf{x}')) + \text{cov}(\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}')) = \eta \exp \left( - \sum_{i=1}^n \frac{(x_i - x'_i)^2}{\theta_i^2} \right) + \sigma^2 \delta(\mathbf{x}, \mathbf{x}'), \quad (1)$$

где  $a = \{\eta, \theta_1, \dots, \theta_n\}$  — набор гиперпараметров ковариационной функции,  $\delta(\mathbf{x}, \mathbf{x}')$  — дельта-функция. Нам нужно оценить гиперпараметры ковариационной функции  $a$  и дисперсию шума  $\sigma^2$ . Будем для этого использовать метод максимального правдоподобия [2, 13]. Правдоподобие данных имеет вид:

$$\log p(\mathbf{y} | X, \sigma^2, a) = - \frac{\mathbf{y}^T K^{-1} \mathbf{y}}{2} - \frac{1}{2} \log |K| - \frac{N}{2} \log 2\pi, \quad (2)$$

где  $|K|$  — детерминант матрицы  $K$ . Максимизируя правдоподобие  $\log p(\mathbf{y} | X, \sigma^2, a)$  по гиперпараметрам ковариационной функции  $a$  и дисперсии шума  $\sigma^2$ :

$$\log p(\mathbf{y} | X, \sigma^2, a) \rightarrow \max_{a, \sigma^2},$$

получаем гиперпараметры максимального правдоподобия  $\hat{a}$ ,  $\hat{\sigma}^2$ . Полученные гиперпараметры будем использовать для оценки математического ожидания и ковариационной матрицы новых точек.

### Регрессия на основе гауссовских процессов для разноточных данных

Пусть теперь верна следующая вероятностная модель [6]:

$$\begin{aligned} y_l(\mathbf{x}) &= f_l(\mathbf{x}) + \varepsilon_l(\mathbf{x}), \\ y_d(\mathbf{x}) &= f_d(\mathbf{x}) + \varepsilon_d(\mathbf{x}), \\ y_h(\mathbf{x}) &= \rho y_l(\mathbf{x}) + y_d(\mathbf{x}), \end{aligned}$$

здесь  $f_l(\mathbf{x})$ ,  $f_d(\mathbf{x})$  — независимые гауссовские процессы с нулевым математическим ожиданием и ковариационными функциями  $k_l(\mathbf{x}, \mathbf{x}')$  и  $k_d(\mathbf{x}, \mathbf{x}')$  соответственно,  $\varepsilon_l(\mathbf{x})$ ,  $\varepsilon_d(\mathbf{x})$  — нормальные случайные величины с нулевым математическим ожиданием и дисперсиями соответственно  $\sigma_l^2$  и  $\sigma_d^2$ . Введем обозначения:

$$X = \begin{pmatrix} X_l \\ X_h \end{pmatrix}, \mathbf{y} = \begin{pmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{pmatrix}.$$

Тогда апостериорное математическое ожидание для новых точек  $X^*$  имеет вид:

$$\mathbb{E} \hat{y}(X^*) = K(X^*, X) K^{-1} \mathbf{y},$$

где

$$K(X^*, X) = \begin{pmatrix} \rho K_l(X^*, X_l) \\ \rho^2 K_l(X^*, X_h) + K_d(X^*, X_h) \end{pmatrix},$$

$$K(X, X) = \begin{pmatrix} K_l(X_l, X_l) & \rho K_l(X_l, X_h) \\ \rho K_l(X_h, X_l) & \rho^2 K_l(X_h, X_h) + K_d(X_h, X_h) \end{pmatrix},$$

где  $K_l(X_a, X_b)$ ,  $K_d(X_a, X_b)$  — матрицы попарных ковариаций гауссовских процессов  $y_l(\mathbf{x})$  и  $y_d(\mathbf{x})$  соответственно для точек из двух выборок  $X_a$  и  $X_b$ . Выражение для апостериорной матрицы ковариации имеет вид:

$$\begin{aligned} \mathbb{V}[X^*] &= \rho^2 K_l(X^*, X^*) + K_d(X^*, X^*) - \\ &K(X^*, X) K^{-1} (K(X^*, X))^T, \end{aligned}$$

Так же как и в предыдущем разделе, мы считаем, что ковариационные функции гауссовских процессов имеют вид (1). Для оценки гиперпараметров ковариационных функций процессов  $f_l(\mathbf{x})$  и  $f_d(\mathbf{x})$  используется следующий алгоритм:

- Оценить гиперпараметры ковариационной функции  $k_l(\mathbf{x}, \mathbf{x}')$ , применив алгоритм, описанный в разделе 1, для выборки  $D_l$ .
- Подсчитать математическое ожидание гауссовского процесса  $f_l(\mathbf{x})$  в точках  $X_h$ .
- Оценить параметры ковариационной функции процесса  $f_d(\mathbf{x})$  и параметр  $\rho$ , максимизируя правдоподобие (2) для  $\mathbf{y} = \mathbf{y}_d = \mathbf{y}_h - \rho \hat{f}_l(X_h)$  и  $k(\mathbf{x}, \mathbf{x}') = k_d(\mathbf{x}, \mathbf{x}')$ .

### Разреженные гауссовские процессы для разноточных данных

Для использования предложенного подхода к оценке математического ожидания гауссовского процесса необходимо обратить матрицу ковариации для обучающей выборки размером  $N \times N$ , где  $N = N_h + N_l$  — размер выборки. Трудоемкость обращения ковариационной матрицы —  $O(N^3)$ . Следовательно, уже при  $N$  порядка нескольких тысяч нельзя эффективно получить оценку апостериорного математического ожидания для новой точки. Для того, что бы ускорить подсчет этой оценки, предлагается применить подход, использующий приближения ковариационных матриц  $K(X^*, X)$ ,  $K$ ,  $K(X^*, X^*)$ . В работе предложен подход, обобщающий результат Фостера [8, 3] на случай гауссовских процессов для разноточных данных.

Пусть задана выборка такого размера, для которого мы можем эффективно обратить матрицу ковариации. Обозначим эту выборку как

$$X^1 = \begin{pmatrix} X_l^1 \\ X_h^1 \end{pmatrix}, \mathbf{y}^1 = \begin{pmatrix} \mathbf{y}_l(X_l^1) \\ \mathbf{y}_h(X_h^1) \end{pmatrix}.$$

Размер такой выборки  $N_1 = N_h^1 + N_l^1$ . Тогда, используя матрицы

$$K_{11} = \begin{pmatrix} K_l(X_l^1, X_l^1) & \rho K_l(X_l^1, X_h^1) \\ \rho K_l(X_h^1, X_l^1) & \rho^2 K_l(X_h^1, X_h^1) + K_d(X_h^1, X_h^1) \end{pmatrix},$$

$$K_1 = \begin{pmatrix} K_l(X_l^1, X_l^1) & \rho K_l(X_l^1, X_h^1) \\ \rho K_l(X_h^1, X_l^1) & \rho^2 K_l(X_h^1, X_h^1) + K_d(X_h^1, X_h^1) \end{pmatrix},$$

$$K_1^* = \begin{pmatrix} \rho K_l(X^*, X_l^1) \\ \rho^2 K_l(X^*, X_h^1) + K_d(X^*, X_h^1) \end{pmatrix},$$

получаем следующие приближения для матриц  $K(X^*, X)$ ,  $K$ ,  $K(X^*, K^*)$ , используя аппроксимация Нёстрема:

$$\hat{K}(X^*, X) = K_1^* K_{11}^{-1} K_1,$$

$$\hat{K} = (K_1)^T K_{11}^{-1} K_1,$$

$$\hat{K}(X^*, X^*) = K_1^* K_{11}^{-1} (K_1^*)^T.$$

Введем обозначение:

$$R = \begin{pmatrix} \frac{1}{\sqrt{\mu_l \sigma_l^2}} I_{N_l} & 0 \\ 0 & \frac{1}{\sqrt{\rho^2 \mu_h \sigma_h^2 + \mu_d \sigma_d^2}} I_{N_h} \end{pmatrix},$$

и  $\tilde{K}_1 = \hat{K}_1 R$ .

Тогда приближение апостериорного математического ожидания имеет вид:

$$\mathbb{E} \hat{y}_h(X^*) \approx \hat{K}_1^* (\hat{K}_{11} + \tilde{K}_1^T \tilde{K}_1)^{-1} \tilde{K}_1^T R y = \quad (3)$$

$$\hat{K}_1^* V_{11} (I + V^T V)^{-1} V^T y, \quad (4)$$

где  $V_{11}$  — разложение Холецкого матрицы  $K_{11}$ ,  $V_1 = \tilde{K}_1 V_{11}^{-T}$ . Приближение апостериорной ковариационной матрицы в точках  $X^*$  имеет вид:

$$\mathbb{V}[X^*] \approx \hat{K}_1^* V_{11} (I + V^T V)^{-1} V_{11}^{-1} (\hat{K}_1^*)^T + (\rho^2 \sigma_l^2 + \sigma_d^2) I_{N^*}.$$

Такой подход позволяет сократить затраты на оценку матрицы  $K(X, X)$  с  $O(N^3)$  до  $O(NN_1^2)$ , где  $N = N_h + N_l$ ,  $N_1 = N_h^1 + N_l^1$ .

## Вычислительный эксперимент

### Модельные данные

Пусть точная функция — функция Растригина:

$$f_h(\mathbf{x}) = 20 + \sum_{i=1}^2 (x_i^2 - 10 \cos(2\pi x_i)),$$

а грубая функция не очень сильно от нее отличается:

$$f_l(\mathbf{x}) = f_h(\mathbf{x}) + 0.2 \sum_{i=1}^2 (\tilde{x}_i + 1)^2,$$

здесь  $\tilde{x}_i = \frac{x_i}{10.28} + 0.5$ .

На рисунке 1 приведено качество работы двух алгоритмов, на рисунке 2 приведено время их работы. В обоих случаях используется выборка точных данных размером 100, по оси абсцисс приведен размер выборки грубых данных. Видно, что с увеличением выборки грубых данных время работы гауссовских процессов для разноточных данных **Full sample VFGP** увеличивается как куб от размера выборки, в то время как для разреженных гауссовских процессов для разноточных данных **Sparse VFGP** время обучения почти не меняется, при этом качество аппроксимации существенно улучшается. Так же на рисунках для сравнения приведено время и качество работы алгоритма для размера выборки грубых данных равного 1000 — **Small sample VFGP**.

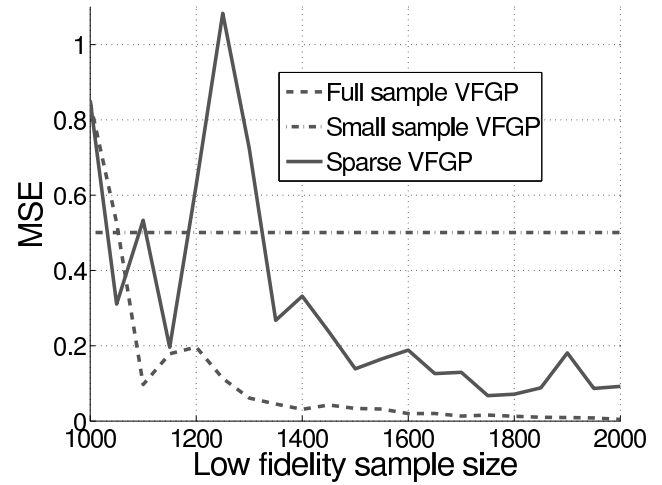


Рис. 1. Сравнение качества работы разреженных и не разреженных гауссовских процессов для разноточных данных

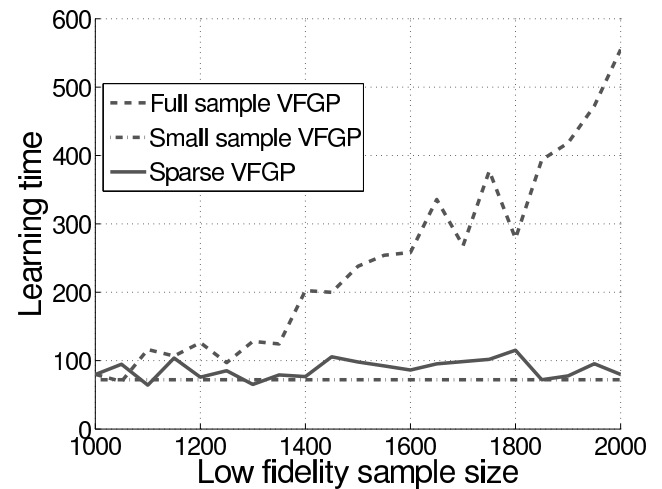


Рис. 2. Сравнение времени работы разреженных и не разреженных гауссовских процессов для разноточных данных

	GP	SGP	VFGP	SVFGP
$C_l$	0.1671	0.0826	0.1284	0.0788
$C_d$	0.0114	0.0068	0.0084	0.0050

Таблица 1. Среднеквадратичная ошибка

## Реальные данные

Рассматривается реальная задача прогнозирования аэродинамических коэффициентов подъемной силы  $C_L$  и сопротивления  $C_D$  профиля крыла самолета. И грубая, и точная модели в этом случае — решение уравнения Эйлера [9], причем для грубой модели используется менее подробная сетка. Получение одного значения целевых функций для точной функции занимает 10–15 минут, а для грубой — 2–3 минуты.

Выборка состояла из 425 профилей крыла для точной функции и 200 профилей крыла для грубой функции. Выборка значений точной функции разбивалась на обучающую и контрольную, в обучающую выборку входило 350 объектов, а в тестовую — 75 объектов, при этом для обучения не разреженных моделей использовалось только 100 объектов из обучающей выборки точных данных. Так же для обучения использовалась полная выборка значений грубой функции. Отдельно строились аппроксимации для коэффициента подъемной силы и сопротивления. В качестве признаков были взяты: параметризация крыла самолета из 9 компонент, параметризующих профиль крыла [5], и угол атаки  $\alpha$ .

Проводилось сравнение четырех алгоритмов:

- **GP** — регрессии на основе гауссовских процессов,
- **SGP** — регрессии на основе разреженных гауссовских процессов,
- **VFGP** — регрессии на основе гауссовских процессов для разноточных данных (variable fidelity),
- **SVFGP** — регрессии на основе разреженных гауссовских процессов для разноточных данных.

Полученные среднеквадратичные ошибки для десятикратного скользящего контроля приведены в таблице 1. Видно, что использование разреженных гауссовских процессов и грубых данных позволяет улучшить качество полученной модели более чем в два раза.

## Выводы

В работе исследован ряд подходов к решению задачи консолидации разноточных данных. На реальной задаче аппроксимации аэродинамических коэффициентов крыла самолета показано, что использование разреженных гауссовских процессов для разноточных данных позволяет существенно улучшить качество полученной аппроксимации.

## Литература

- [1] M.A. Álvarez and N.D. Lawrence. Computationally efficient convolved multiple output gaussian processes. *Journal of Machine Learning Research*, 12:1425–1466, 2011.
- [2] C.M. Bishop. *Pattern recognition and machine learning*. Springer New York, 2006.
- [3] E. Burnaev and Yu. Yanovich. Sparse gaussian processes. In *Conference MIPT-54*, 2011.
- [4] E. Burnaev, A. Zaytsev, M. Panov, P. Prihodko, and Yu. Yanovich. Modeling of nonstationary covariance function of gaussian process using decomposition in dictionary of nonlinear functions. In *Information Technologies and Systems-2011*, Moscow, October 2–7 2011.
- [5] П.Д. Ерофеев, П.В. Приходько. Применение техники эффективного снижения размерности в задаче оптимизации крыла самолета. In *Conference MIPT-54*, 2011.
- [6] A.I.J. Forrester, A. Sóbester, and A.J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 463(2088):3251–3269, 2007.
- [7] A.I.J. Forrester, A. Sóbester, and A.J. Keane. *Engineering design via surrogate modelling: a practical guide*. Progress in astronautics and aeronautics. J. Wiley, 2008.
- [8] L. Foster, A. Waagen, N. Aijaz, M. Hurley, A. Luis, J. Rinsky, C. Satyavolu, M.J. Way, P. Gazis, and A. Srivastava. Stable and efficient gaussian process calculations. *The Journal of Machine Learning Research*, 10:857–882, 2009.
- [9] , A. Jameson, W. Schmidt, E. Turkel et al. Numerical solutions of the Euler equations by finite volume methods using Runge-Kutta time-stepping schemes. *AIAA paper*, 1981
- [10] Z. Han, S. Görtz, and R. Hain. A variable-fidelity modeling method for aero-loads prediction. *New Results in Numerical and Experimental Fluid Mechanics*, 7:17–25, 2010.
- [11] M.C. Kennedy and A. O’Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.
- [12] P.Z.G. Qian and C.F.J. Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204, 2008.
- [13] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006.
- [14] A. A. Zaytsev, E. V. Burnaev, and E. R. Kapushev. The data fusion problem. In *Conference MIPT-54*, 2011.