Информационные процессы, Том 15, № 1, 2015, стр. 97–109

Суррогатное моделирование разноточных данных в случае выборок большого размера[1]

Е.В. Бурнаев*,**,***, А.А. Зайцев*,**

*Институт проблем передачи информации, Российская академия наук, Москва, Россия
**ООО Датадванс, Москва, Россия
***Лаборатория предсказательного моделирования и анализа данных,
Московский Физико-Технический Институт (Государственный Университет), Москва, Россия
e-mail: burnaev@iitp.ru

Поступила в редколлегию 27.03.2015

Аннотация—В работе рассматривается задача построения суррогатной модели в случае наличия данных низкой и высокой точности. Данные низкой точности могут быть получены, например, в результате проведения компьютерного моделирования, а данные высокой точности – в результаты проведения экспериментов в аэродинамической трубе. Для моделирования разноточных данных удобной оказывается модель регрессии на основе гауссовских процессов, с помощью которой можно эффективно восстанавливать нелинейные зависимости и оценивать точность прогноза зависимости в заданной точке. Однако, для выборок с размерами больше нескольких тысяч точек непосредственное применение регрессии на основе гауссовских процессов становится невозможным из-за высокой вычислительной сложности алгоритма. В данной работе предложены алгоритмы, которые позволяют обрабатывать выборки разноточных данных с помощью аппарата регрессии на основе гауссовских процессов даже в случае выборок большого размера. Приводятся примеры применения разработанных алгоритмов для решения задач суррогатного моделирования сложных инженерных конструкций.

Ключевые слова: разноточные данные, оценка неопределенности, гауссовские процессы, аппроксимация ковариационной матрицы, кокригинг

1. ВВЕДЕНИЕ

В инженерной практике весьма часто возникает задача консолидации разноточных данных: нужно построить регрессионную моделей в случае, если помимо выборки значений точной функции задана вторая выборка бóльшего размера, содержащая значения грубой функции (некоторого приближения точной функции) [1].

Например, точная функция – значения подъемной силы крыла самолета, измеренные в аэродинамической трубе, а грубая функция – значения подъемной силы крыла самолета, полученные с помощью численного моделирования. Эксперименты в аэродинамической трубе позволяют получать точные значения подъемной силы, но при этом затратны по времени и ресурсам в отличии от виртуальных экспериментов, проводимых на основе компьтерного моделирования. Таким образом, если мы строим регрессионную модель зависимости подъемной силы крыла самолета от геометрических параметров крыла, мы вынуждены работать с небольшой выборкой данных, полученной с помощью аэродинамической трубы, и большой выборкой данных, полученной с помощью менее точного численного моделирования.

[1] Исследование выполнено в ИППИ РАН исключительно за счет гранта Российского научного фонда (проект № 14-50-00150)

Для моделирования разноточных данных удобной оказывается модель регрессии на основе гауссовских процессов [1–3], с помощью которой можно эффективно восстанавливать нелинейные зависимости и оценивать точность прогноза зависимости в заданной точке [4–6].

Размер выборки, которая может использоваться для построения суррогатной модели с использованием регрессии на основе гауссовских процессов, ограничен несколькими тысячами точек, так как в процессе оценки параметров регрессии необходимо обращать матрицу ковариаций точек выборки [7]. Поэтому если имеется выборка данных только одной точности, но большого размера, то для построения регрессии на основе гауссовских процессов используют специальную аппроксимацию исходной ковариационной матрицы [7–9] (аппроксимация Нистрема [10]) на основе некоторого подмножества базовых точек, которая позволяет существенно сократить сложность вычислений. Однако для построения регрессии на основе гауссовских процессов по разноточным данным до настоящего времени не было предложено способа работы с выборками размера больше нескольких тысяч точек. В то же время, для разноточных данных выборки большого размера встречаются чаще, поскольку "стоимость" вычисления одного значения грубой функции обычно значительно ниже "стоимости" вычисления одного значения точной функции [4], и по этой причине выборка данных низкой точности может иметь большой размер.

В данной работе рассматривается подход, позволяющий применять регрессию на основе гауссовских процессов для разноточных данных в случае выборок большого размера. Основная идея подхода заключается в использовании подвыборки исходной выборки точек для аппроксимации ковариационной матрицы выборки [9]. В работе получены: оценки апостериорного среднего регрессии, которое используется в качестве прогноза значения точной функции, и апостериорной дисперсии регрессии, которая используется для оценки неопределенности прогноза значения точной функции; оценки вычислительной сложности предложенного алгоритма.

Работа содержит следующие разделы: в разделе 2 описывается регрессия на основе гауссовских процессов; в разделе 3 показано, как использовать регрессию на основе гауссовских процессов для моделирования разноточных данных; в разделе 4 описан предложенный алгоритм для построения суррогатной модели на основе разноточных данных в случае выборок большого размера; раздел 5 содержит результаты экспериментов на искусственных и реальных данных; выводы представлены в разделе 6.

2. РЕГРЕССИЯ НА ОСНОВЕ ГАУССОВСКИХ ПРОЦЕССОВ

Рассмотрим обучающую выборку $D = (X, \mathbf{y}) = \{\mathbf{x}_i, y_i = y(\mathbf{x}_i)\}_{i=1}^n$, где точки $\mathbf{x} \in \mathbb{X} \subseteq \mathbb{R}^d$ и значения функции $y(\mathbf{x}) \in \mathbb{R}$. Мы предполагаем, что $y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon$, где функция $f(\mathbf{x})$ – реализация гауссовского процесса, а ε – гауссовский белый шум с дисперсией σ^2. Необходимо построить суррогатную модель для целевой функции $f(\mathbf{x})$.

Среднее значение и ковариционная функция гауссовского процесса

$$k(\mathbf{x}, \mathbf{x}') = \operatorname{cov}(f(\mathbf{x}), f(\mathbf{x}')) = \mathbb{E}\left(f(\mathbf{x}) - \mathbb{E}(f(\mathbf{x}))\right)\left(f(\mathbf{x}') - \mathbb{E}(f(\mathbf{x}'))\right)$$

полностью определяют гауссовский процесс $f(\mathbf{x})$. Для упрощения изложения будем считать его среднее значение равным нулю. Мы предполагаем, что ковариационная функция лежит в параметрическом семействе $\{k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}'), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, то есть $k(\mathbf{x}, \mathbf{x}') = k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}')$ для некоторого $\boldsymbol{\theta} \in \Theta$. Тогда $y(\mathbf{x})$ будет гауссовским процессом с нулевым средним и ковариационной функцией $\operatorname{cov}(y(\mathbf{x}), y(\mathbf{x}')) = k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') + \sigma^2\delta(\mathbf{x} - \mathbf{x}')$, где $\delta(\mathbf{x} - \mathbf{x}')$ – дельта функция. Широко используемым классом ковариационных функций является, например, квадратичная экспоненциальная ковариационная функция [3] $k_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp\left(-\sum_{k=1}^d \theta_k^2 (x_k - x'_k)^2\right)$.

Параметры ковариационной функции $\boldsymbol{\theta}$ и дисперсия шума σ^2 задают модель регрессии. Мы используем метод максимального правдоподобия для оценки параметров $\boldsymbol{\theta}$ и σ^2 [3]:

$$\log p(\mathbf{y}|X,\boldsymbol{\theta},\sigma^2) = -\frac{1}{2}\left(n\log 2\pi + \log|K| + \mathbf{y}^T K^{-1}\mathbf{y}\right) \to \max_{\boldsymbol{\theta},\sigma^2}, \tag{1}$$

где $K = \{k_{\boldsymbol{\theta}}(\mathbf{x}_i,\mathbf{x}_j) + \sigma^2\delta(\mathbf{x}_i - \mathbf{x}_j)\}_{i,j=1}^n$ – матрица ковариаций между значениями целевой функции $\mathbf{y}(X)$ в обучающей выборки, и $|K|$ – детерминант матрицы K. В регрессии на основе гауссовских процессов σ^2 играет роль параметра регуляризации для ковариационной матрицы $\{k_{\boldsymbol{\theta}}(\mathbf{x}_i,\mathbf{x}_j)\}_{i,j=1}^n$ значений $f(X)$. Теоретические результаты, полученные в работах [11], и более прикладных исследования [12] показывают, что таким образом полученные оценки параметров $\hat{\boldsymbol{\theta}}$ точны даже если размер выборки невелик, а модель, определяемая ковариационной функцией, неправильно специфицирована.

Используя оценки параметров $\boldsymbol{\theta}$ и σ^2 можно вычислить апостериорное среднее и дисперсию $y(\mathbf{x})$ в новых точках, которые используются для прогноза значений функции и оценки неопределенности прогноза. Апостериорное среднее $\mathbb{E}(\mathbf{y}(X^*)|\mathbf{y}(X))$ в новых точках $X^* = \{\mathbf{x}_i^*\}_{i=1}^{n^*}$ имеет вид

$$\hat{\mathbf{y}}(X^*) = K(X^*,X)K^{-1}\mathbf{y}, \tag{2}$$

где $K(X^*,X) = \{k(\mathbf{x}_i^*,\mathbf{x}_j)\}_{i=1,\ldots,n^*,j=1,\ldots,n}$ – ковариации между $\mathbf{y}(X^*)$ в новых точках и значениями $\mathbf{y}(X)$ в точках из обучающей выборки. Апостериорная ковариационная матрица $\mathbb{V}(X^*) = \mathbb{E}\big[(\mathbf{y}(X^*) - \mathbb{E}\mathbf{y}(X^*))^T(\mathbf{y}(X^*) - \mathbb{E}\mathbf{y}(X^*)) \,|\, \mathbf{y}(X)\big]$ в новых точках имеет вид

$$\mathbb{V}(X^*) = K(X^*,X^*) - K(X^*,X)K^{-1}(K(X,X^*))^T, \tag{3}$$

где $K(X^*,X^*) = \{k(\mathbf{x}_i^*,\mathbf{x}_j^*) + \sigma^2\delta(\mathbf{x}_i^* - \mathbf{x}_j^*)\}_{i,j=1}^{n^*}$ – матрица ковариаций между значениями $\mathbf{y}(X^*)$.

3. РЕГРЕССИЯ НА ОСНОВЕ ГАУССОВСКИХ ПРОЦЕССОВ ДЛЯ РАЗНОТОЧНЫХ ДАННЫХ

Теперь рассмотрим случай разноточных данных: пусть задана выборка значений грубой функции $D_l = (X_l,\mathbf{y}_l) = \left\{\mathbf{x}_i^l, y_l(\mathbf{x}_i^l)\right\}_{i=1}^{n_l}$ и выборка значений точной функции $D_h = (X_h,\mathbf{y}_h) = \left\{\mathbf{x}_i^h, y_h(\mathbf{x}_i^h)\right\}_{i=1}^{n_h}$ с $\mathbf{x}_i^l, \mathbf{x}_i^h \in \mathbb{R}^d$, $y_l(\mathbf{x}), y_h(\mathbf{x}) \in \mathbb{R}$. Грубая функция $y_l(\mathbf{x})$ и точная функция $y_h(\mathbf{x})$ моделируют одно и то же физическое явление, но с разной точностью.

Используя выборки значений грубой и точной функции, необходимо построить как можно более точную суррогатную модель $\hat{y}_h(\mathbf{x}) \approx y_h(\mathbf{x})$ точной функции. Кроме того, мы хотим получить оценку неопределенности прогноза значений точной функции в новых точках.

Для моделирования данных разной точности необходима специальная модель. Мы используем широко распространенную модель кокригинга [4]:

$$y_l(\mathbf{x}) = f_l(\mathbf{x}) + \varepsilon_l,\ y_h(\mathbf{x}) = \rho y_l(\mathbf{x}) + y_d(\mathbf{x}),$$

где $y_d(\mathbf{x}) = f_d(\mathbf{x}) + \varepsilon_d$, $f_l(\mathbf{x})$, $f_d(\mathbf{x})$ – независимые гауссовские процессы с нулевыми средними и ковариационными функциями $k_l(\mathbf{x},\mathbf{x}')$ и $k_d(\mathbf{x},\mathbf{x}')$ соответственно, и ε_l, ε_d – гауссовский белый шум с дисперсиями σ_l^2 и σ_d^2, соответственно. Обозначим $X = \begin{pmatrix} X_l \\ X_h \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{pmatrix}$. Тогда апостериорное среднее для значений точной функции в новых точках имеет вид

$$\hat{\mathbf{y}}_h(X^*) = K(X^*,X)K^{-1}\mathbf{y}, \tag{4}$$

где

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

$K_l(X_a, X_b)$, $K_d(X_a, X_b)$ – матрицы попарных ковариаций гауссовских процессов $y_l(\mathbf{x})$ и $y_d(\mathbf{x})$ в точках из некоторых множеств X_a and X_b соответственно. Апостериорная ковариационная матрица имеет вид:

$$\mathbb{V}(X^*) = \rho^2 K_l(X^*, X^*) + K_d(X^*, X^*) - K(X^*, X) K^{-1} (K(X^*, X))^T . \tag{5}$$

Для оценки параметров ковариационных функций гауссовских процессов $f_l(\mathbf{x})$ и $f_d(\mathbf{x})$ применяется следующий алгоритм [1]:

1. Оценить параметры ковариационной функции $k_l(\mathbf{x}, \mathbf{x})$, используя алгоритм для обычной регрессии на основе гауссовских процессов, описанный в разделе 2 с выборкой $D = D_l$,
2. Подсчитать оценки апостериорного среднего $\hat{y}_l(\mathbf{x})$ для гауссовского процесса $y_l(\mathbf{x})$ и $\mathbf{x} \in X_h$,
3. Оценить параметры гауссовского процесса $y_d(\mathbf{x})$ с ковариационной функцией $k_d(\mathbf{x}, \mathbf{x}')$ и параметр ρ, максимизируя правдоподобие (1) с $D = D_{\text{diff}} = (X_h, \mathbf{y}_d = \mathbf{y}_h - \rho \hat{\mathbf{y}}_l(X_h))$ и $k(\mathbf{x}, \mathbf{x}') = k_d(\mathbf{x}, \mathbf{x}')$.

4. РАЗРЕЖЕННАЯ РЕГРЕССИЯ НА ОСНОВЕ ГАУССОВСКИХ ПРОЦЕССОВ ДЛЯ РАЗНОТОЧНЫХ ДАННЫХ

Эффективное использование регрессии на основе гауссовских процессов для разноточных данных возможно только для выборок с размером не больше нескольких тысяч точек: в процессе обучения модели и ее использования приходится обращать ковариационную матрицу размера $n \times n$, где $n = n_h + n_l$. Вычислительная сложность такой процедуры составляет $O(n^3)$.

Мы предлагаем новый подход для регрессии на основе гауссовских процессов в случае выборок разноточных данных большого размера. Подход основан на использовании аппроксимации Нистрема матриц $K(X^*, X)$, K и $K(X^*, X^*)$ с использованием подвыборки базовых точек исходной выборки. Представленные результаты являются обобщением результатов работы [9] на случай разноточных данных.

Зададим подвыборку $D_1 = (X^1, \mathbf{y}^1)$, $X^1 = \begin{pmatrix} X_l^1 \\ X_h^1 \end{pmatrix}, \mathbf{y}^1 = \begin{pmatrix} \mathbf{y}_l(X_l^1) \\ \mathbf{y}_h(X_h^1) \end{pmatrix}$ базовых точек из исходной выборки такого размера $n_1 = n_h^1 + n_l^1$, что для заданной подвыборки имеющиеся вычислительные ресурсы позволяют обратить соответствующие ковариационные матрицы и оценить параметры гауссовскго процесса за разумное время. Достаточно надежный метод задания подвыборки базовых точек состоит в случайном выборе без повторения точек из исходной выборки.

Тогда используя подвыборку базовых точек и положив

$$K_{11} = \begin{pmatrix} K_l(X_l^1, X_l^1) & \rho K_l(X_l^1, X_h^1) \\ \rho K_l(X_h^1, X_l^1) & \rho^2 K_l(X_h^1, X_h^1) + K_d(X_h^1, X_h^1) \end{pmatrix},$$
$$K_1 = \begin{pmatrix} K_l(X_l^1, X_l) & \rho K_l(X_l^1, X_h) \\ \rho K_l(X_h^1, X_l) & \rho^2 K_l(X_h^1, X_h) + K_d(X_h^1, X_h) \end{pmatrix}, K_1^* = \begin{pmatrix} \rho K_l(X^*, X_l^1) \\ \rho^2 K_l(X^*, X_h^1) + K_d(X^*, X_h^1) \end{pmatrix}$$

для новых точек $X^* = \{\mathbf{x}_i^*\}_{i=1}^{n^*}$, получаем аппроксимации матриц $K(X^*, X)$, K и $K(X^*, X^*)$ соответственно:

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

Определим

$$R = \begin{pmatrix} \frac{1}{\sigma_l} I_{n_l} & 0 \\ 0 & \frac{1}{\sqrt{\rho^2\sigma_l^2 + \sigma_d^2}} I_{n_h} \end{pmatrix},$$

где I_k – единичная матрица размера k, $C_1 = RK_1$, и $V = C_1 V_{11}^{-T}$, V_{11} – разложение Холецкого [13] матрицы K_{11}.

Утверждение 1. Для апостериорного среднего аппроксимация Нистрема имеет вид:

$$\hat{\mathbf{y}}_h(X^*) \approx K_1^* V_{11} (I_{n_1} + V^T V)^{-1} V^T \mathbf{y}, \tag{6}$$

Доказательство. Действительно,

$$\begin{aligned}
\hat{\mathbf{y}}_h(\mathbf{x}^*) &\approx K_1^* K_{11}^{-1} K_1^T (K_1 K_{11}^{-1} K_1^T + R^{-2})^{-1} \mathbf{y} = K_1^* K_{11}^{-1} K_1^T R (R K_1 K_{11}^{-1} K_1^T R + I_n)^{-1} R\mathbf{y} = \\
&= K_1^* K_{11}^{-1} C_1^T (C_1 K_{11}^{-1} C_1^T + I_n)^{-1} R\mathbf{y} = K_1^* K_{11}^{-1} (C_1^T C_1 K_{11}^{-1} + I_{n_1})^{-1} C_1^T R\mathbf{y} = \\
&= K_1^* (C_1^T C_1 + K_{11})^{-1} C_1^T R\mathbf{y} = K_1^* (C_1^T C_1 + V_{11}^T V_{11})^{-1} C_1^T R\mathbf{y} = \\
&= K_1^* V_{11}^{-1} (V_{11}^{-T} C_1^T C_1 V_{11}^{-1} + I_{n_1})^{-1} V_{11}^{-T} C_1^T R\mathbf{y} = K_1^* V_{11}^{-1} (V^T V + I_{n_1})^{-1} V^T R\mathbf{y}.
\end{aligned}$$

Для оценки неопределенности прогноза результат будет зависеть от используемого приближения ковариационных матриц. Рассмотрим три возможных варианта приближения, приведенных в таблице 1.

Вариант приближения	$k(\mathbf{x}^*, \mathbf{x}^*)$	$k(\mathbf{x}^*, X)$	$k(X, X)$
1	$K_1^* K_{11}^{-1} K_1^{*T}$	$K_1^* K_{11}^{-1} K_1^T$	$R^{-2} + K_1 K_{11}^{-1} K_1^T$
2	$k(\mathbf{x}^*, \mathbf{x}^*)$	$K_1^* K_{11}^{-1} K_1^T$	$K_1 K_{11}^{-1} K_1^T$
3	$k(\mathbf{x}^*, \mathbf{x}^*)$	$K_1^* K_{11}^{-1} K_1^T$	$R^{-2} + K_1 K_{11}^{-1} K_1^T$

Таблица 1. Аппроксимация ковариационных матриц для разных оценок апостериорной дисперсии прогноза

Утверждение 2. Для апостериорной дисперсии в случае использования вариантов приближений 1, 2 и 3 из таблицы 1 аппроксимации Нистрема будут иметь следующий вид:

$$\hat{\sigma}_1^2(\mathbf{x}^*) = K_1^* V_{11}^{-1} (I + V^T V)^{-1} V_{11}^{-T} K_1^{*T}, \tag{7}$$

$$\hat{\sigma}_2^2(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - K_1^* V_{11}^{-1} V_{11}^{-T} K_1^{*T}, \tag{8}$$

$$\hat{\sigma}_3^2(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - K_1^* V_{11}^{-1} (I + V^T V)^{-1} (V^T V) V_{11}^{-T} K_1^{*T}. \tag{9}$$

Доказательство. Используя первый набор приближений, получаем следующее выражение для дисперсии:

$$\begin{aligned}
\hat{\sigma}_1^2(\mathbf{x}^*) &\approx K_1^* K_{11}^{-1} K_1^{*T} - K_1^* K_{11}^{-1} K_1^T (R^{-2} + K_1 K_{11}^{-1} K_1^T)^{-1} K_1 K_{11}^{-1} K_1^{*T} = \\
&= K_1^* (K_{11}^{-1} - K_{11}^{-1} K_1^T (R^{-2} + K_1 K_{11}^{-1} K_1^T)^{-1} K_1 K_{11}^{-1}) K_1^{*T} \\
&= K_1^* (K_{11} + K_1^T R^2 K_1)^{-1} K_1^{*T} = \\
&= K_1^* (V_{11}^T V_{11} + C_1^T C_1)^{-1} K_1^{*T} = \\
&= K_1^* V_{11}^{-1} (I + V^T V)^{-1} V_{11}^{-T} K_1^{*T}.
\end{aligned}$$

Приведенное выше выражение сходно с тем, которое получается в [9].

Получим теперь результат с использованием второго варианта приближений для ковариационных матриц:

$$\begin{aligned}\hat{\sigma}_2^2(\mathbf{x}^*) &= k(\mathbf{x}^*,\mathbf{x}^*) - k(\mathbf{x}^*,X)k(X,X)^{-1}k(\mathbf{x}^*,X)^T \approx \\ &\approx k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} K_1^T (K_1 K_{11}^{-1} K_1^T)^{-1} K_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} K_1^{*T}.\end{aligned}$$

Полученная оценка дисперсии совпадает с оценкой дисперсии в случае, если мы используем обычную регрессию на основе гауссовских процессов и выборку D_1. Поэтому ясно, что она обладает достаточно хорошими численными свойствами, но никак не учитывает информацию о том, что для вычисления прогноза использовалась дополнительная выборка $D \setminus D_1$.

Получим теперь выражение в случае использования третьего варианта приближений ковариационных матриц:

$$\begin{aligned}\hat{\sigma}_3^2(\mathbf{x}^*) &\approx k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} K_1^T (R^{-2} + K_1 K_{11}^{-1} K_1^T)^{-1} K_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} K_1^T R (I + R K_1 K_{11}^{-1} K_1^T R)^{-1} R K_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} C_1^T (I + C_1 K_{11}^{-1} C_1^T)^{-1} C_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* K_{11}^{-1} (I + C_1^T C_1 K_{11}^{-1})^{-1} C_1^T C_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (K_{11} + C_1^T C_1)^{-1} C_1^T C_1 K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (I + (C_1^T C_1)^{-1} K_{11})^{-1} K_{11}^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (K_{11} + K_{11} (C_1^T C_1)^{-1} K_{11})^{-1} K_1^{*T}.\end{aligned}$$

Преобразуем полученное выражение с учетом введенных ранее обозначений:

$$\begin{aligned}\hat{\sigma}_3^2(\mathbf{x}^*) &\approx k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (K_{11} + K_{11} (C_1^T C_1)^{-1} K_{11})^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (V_{11}^T V_{11} + V_{11}^T V_{11} (C_1^T C_1)^{-1} V_{11}^T V_{11})^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* (V_{11}^T V_{11} + V_{11}^T (V^T V)^{-1} V_{11})^{-1} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* V_{11}^{-1} (I + (V^T V)^{-1})^{-1} V_{11}^{-T} K_1^{*T} = \\ &= k(\mathbf{x}^*,\mathbf{x}^*) - K_1^* V_{11}^{-1} (I + V^T V)^{-1} V^T V V_{11}^{-T} K_1^{*T}.\end{aligned}$$

Отметим, что если мы хотим избежать обращения матрицы $(V^T V)^{-1}$, обусловленность которой мы не можем напрямую контролировать, необходимо использовать выражение, несимметричное относительно входящих в него матриц. Стоит так же отметить, что если матрица $V^T V$ существенно больше по масштабу единичной, то $(I+V^T V)^{-1}(V^T V) \approx I$ и $\hat{\sigma}_3^2(\mathbf{x}^*) \approx \hat{\sigma}_2^2(\mathbf{x}^*)$, то есть эти два варианта оценки неопределенности в таком случае почти совпадают.

Утверждение 3. Вычислительная сложность подсчета в одной точке апостериорного среднего с использованием (6) и апостериорной дисперсии с использованием (7), (8) или (9) равна $O(nn_1^2)$.

Доказательство. Сначала необходимо вычислить матрицы V_{11} и $V = RK_1 V_{11}^{-T}$. Матрица V_{11} имеет размер $n_1 \times n_1$, и чтобы подсчитать ее обратную матрицу нужно $O(n_1^3)$ операций. Подсчет $K_1 V_{11}^{-T}$ требует $O(n_1^2 n)$ операций. Для того, чтобы теперь подсчитать V нужно $O(n_1 n)$ операций, так как матрица R – диагональная.

Для $n^* = 1$ вычисление апостериорного среднего состоит в вычислении $V_{11}(I_{n_1}+V^T V)^{-1} V^T \mathbf{y}$. Мы используем $O(n_1^2 n)$ операций для подсчета $V^T V$. Для того, чтобы обратить $I_{n_1} + V^T V$

нужно $O(n_1^3)$ операций, Чтобы подсчитать $V_{11}(I_{n_1} + V^T V)^{-1} V^T$ необходимо $O(n_1^2 n)$ операций. Наконец, оценка апостериорного среднего требует еще $O(n_1 n)$ операций. Следовательно, трудоемкость подсчета приближения Нистрема апостериорного среднего равна $O(n_1^2 n)$ операций.

Для того, чтобы подсчитать $V_{11}(I_{n_1} + V^T V)^{-1} V_{11}^{-1}$, нужно $O(n_1^2 n)$ операций для вычисления $(I_{n_1} + V^T V)^{-1}$ и еще $O(n_1^3)$ операций для получения окончательного результата. Следовательно, трудоемкость подсчета приближения апостериорной дисперсии с использованием (7) требует $O(n_1^2 n)$ операций. Аналогичным образом получаем трудоемкость для вычисления аппроксимации апостериорной дисперсии с использованием (8) и (9).

Таким образом, необходимо $O(n_1^2 n)$ операций для вычисления требуемых матриц и $O(n_1^2 n)$ операций для вычисления апостериорного среднего и апостериорной дисперсии используя эти матрицы. Следовательно, общая вычислительная сложность равна $O(n_1^2 n)$.

5. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В этом разделе мы рассмотрим решение нескольких искусственных тестовых задач и одной реальной задачи с помощью предложенного подхода к разреженной регрессии на основе гауссовских процессов для разноточных данных (Sparse Variable Fidelity Gaussian Processes Regression – SVFGP). Предложеный подход сравнивается с регрессией на основе гауссовских процессов (Gaussian Processes Regression – GP) для данных одной точности и регрессией на основе гауссовских процессов для разноточных данных (Variable Fidelity Gaussian Processes Regression – VFGP), для которой не используется аппроксимация Нистрема. В экспериментах используется квадратичная экспоненциальная ковариационная функцая [3].

В качестве меры качества полученных суррогатных моделей мы используем ошибку RRMS, которая оценивается с помощью скользящего контроля. Для тестовой выборки $D_{\text{test}} = \{\mathbf{x}_i^{\text{test}}, y_i^{\text{test}} = f_h(\mathbf{x}_i^{\text{test}})\}_{i=1}^{n_t}$ ошибка RRMS суррогатной модели $\hat{y}(\mathbf{x})$ равна

$$RRMS(D_{\text{test}}, \hat{y}) = \frac{\sum_{i=1}^{n_t} (\hat{y}_h(\mathbf{x}_i^{\text{test}}) - y_i^{\text{test}})^2}{\sum_{i=1}^{n_t} (\overline{y} - y_i^{\text{test}})^2},$$

здесь $\overline{y} = \frac{1}{n_t} \sum_{i=1}^{n_t} y_i^{\text{test}}$. Обычно значения ошибки RRMS лежат между 0 и 1. У точных суррогатных моделей значения ошибки RRMS близки к нулю, у неточных суррогатных моделей значения ошибки RRMS близки или превосходят 1.

5.1. Искусственные данные

Для тестирования предложенного нами подхода SVFGP мы используем искусственную функцию с большим количеством локальных особенностей и входной размерностью $d = 5$. Таким образом, для того, чтобы построить точную суррогатную модель, нам нужна выборка большого размера. В качестве точной функции $y_h(\mathbf{x})$ и грубой функции $y_l(\mathbf{x})$ мы используем

$$y_h(\mathbf{x}) = 20 + \sum_{i=1}^{d} (x_i^2 - 10\cos(2\pi x_i)) + \varepsilon_h,\ \mathbf{x} \in [0,1]^d,$$

$$y_l(\mathbf{x}) = y_h(\mathbf{x}) + 0.2 \sum_{i=1}^{d} (x_i + 1)^2 + \varepsilon_l,\ \mathbf{x} \in [0,1]^d.$$

Точная функция была зашумлена гауссовским белым шумом ε_h с дисперсией 0.001, и грубая функция была зашумлена гауссовским белым шумом ε_l с дисперсией 0.002. Мы генерируем точки из гиперкуба $[0,1]^d$ с использованием оптимальных латинских гиперкубов (OLHS, [14]). Размер выборки точных данных равнялся $n_h = 100$, размер подвыборки базовых точек для

SVFGP равнялся $n_l^1 = 1000$ во всех экспериментах в данном разделе. Рассматривались значения размера выборки неточных данных $n_l = 1000, 2000, 3000, 40000, 5000$. Результаты были усреднены по 50 запускам для каждого значения n_l.

Для вычислений использовался персональный компьютер с операционной системой Ubuntu, 4 ядрами Intel-Core $i7$ с тактовой частотой 3.4 ГГц и 8 гигабайтами оперативной памяти. Полученные результаты представлены в таблицах:

– RRMS ошибки для VFGP и SVFGP представлены в таблице 2,
– время обучения суррогатных моделей для VFGP и SVFGP представлены в таблице 3.

Ошибки RRMS для SVFGP сравнимы с ошибками RRMS для VFGP для того же размера обучающей выборки, но время обучения модели для SVFGP существенно меньше, особенно для размеров выборок близких к 5000.

n_l	1000	2000	3000	4000	5000
VFGP	0.0100	0.0086	0.0028	0.0031	0.0024
SVFGP	0.0100	0.0067	0.0049	0.0044	0.0044

Таблица 2. Сравнение ошибок RRMS для VFGP и SVFGP на искусственных данных

n_l	1000	2000	3000	4000	5000
VFGP	23.83	254.4	758.2	2334	4496
SVFGP	23.36	26.07	28.89	29.49	35.33

Таблица 3. Сравнение времени обучения моделей в секундах для VFGP и SVFGP на искусственных данных

5.2. Оценки неопределенности прогноза для искусственных данных

Рассмотрим функцию Экли [15] для трехмерного пространства входов:

$$f(\mathbf{x}) = \exp(-0.2\sqrt{x_1^2 + x_2^2}) + 3(\cos(2x_1) + \sin(2x_2)) + \\ + \exp(-0.2\sqrt{x_2^2 + x_3^2}) + 3(\cos(2x_2) + \sin(2x_3)).$$

Для обучения модели использовались точная и грубая функция, которые отличались дисперсиями шума:

$$y_h(\mathbf{x}) = f(\mathbf{x})(1 + 2\sigma_h^2\varepsilon),$$
$$y_l(\mathbf{x}) = f(\mathbf{x})(1 + 2\sigma_l^2\varepsilon),$$

ε – гауссовский белый шум с единичной дисперсией. Для точной функции дисперсия σ_h^2 равнялась 0.1, а размер выборки равнялся 60. Для грубой функции дисперсия σ_l^2 равнялась 0.4, а размер выборки равнялся 160.

Будем сравнивать качество оценок неопределенности на независимой тестовой выборке. На рисунке 1 изображены функция распределения реальных ошибок прогнозов и функция распределения оценок неопределенностей прогнозов, полученных с помощью формул (7), (8) и (9) соответственно. Видно, что для оценок неопределенности (8) и (9) функции распределения почти совпадают и ближе к истинной функции распределения ошибок, чем функция распределения оценок неопределенностей, полученных с помощью формулы (7), использование которой приводит к занижению значений ошибок.

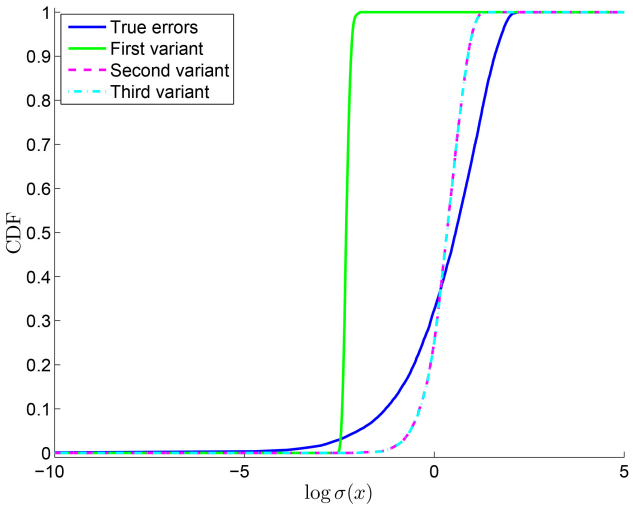

Рис. 1. Функция распределения логарифмов ошибок для реальных ошибок (True errors) и оценок неопределенности с использованием (7) (First variant), (8) (Second variant) и (9) (Third variant)

Численные результаты о качестве оценок ошибок приведены в таблице 4 и позволяют прийти к аналогичным выводам: формулы (8) и (9) обеспечивают более точные оценки неопределенностей с точки зрения корреляции и критерия RRMS, подсчитанного по истинным значениям ошибок и их прогнозам.

Формула для приближения	корреляция	RRMS ошибка
(7)	-0.0274	1.2730
(8)	0.4756	0.6769
(9)	0.4757	0.6769

Таблица 4. Качество оценок неопределенности, полученных с помощью разных приближений

5.3. Задача о вращающемся диске

Вращающийся диск – важная деталь двигателя самолета. Необходимо построить точные модели для прогноза максимального радиального смещения $u_{\max}$ и максимальной нагрузки $s_{\max}$, которые определяют надежность диска [16].

Мы параметризуем геометрию вращающегося диска используя 8 параметров: радиусы r_i, $i = 1, \ldots, 6$, которые определяют, где происходят изменения толщины диска, и значения t_1, t_3, t_5, которые определяют собственно толщину диска. В рассматриваемой задаче мы фиксируем радиусы r_4, r_5 и толщину t_3 вращающего диска, поэтому размерность пространства параметров диска равна 6. Геометрические параметры вращающего диска изображены на рисунке 2.

Мы рассматриваем следующие точные и грубые функции для вычисления целевых значений $u_{\max}$ и $s_{\max}$: в качестве грубой функции мы используем солвер на основе обыкновенных дифференциальных уравнений, реализующий метод Рунге–Кутты [17], в качестве точной функции мы используем солвер на основе метода конечных элементов. Одно вычисление грубой функции требует примерно 0.01 секунду, одно вычисление точной функции требует примерно 300 секунд.

Примеры срезов грубой и точной функций представлены на рисунках 3, 4, 5 для выхода $s_{\max}$. Грубая и точная функции похожи друг на друга, но в некоторых случаях грубая функция не способна моделировать некоторые нелинейные эффекты

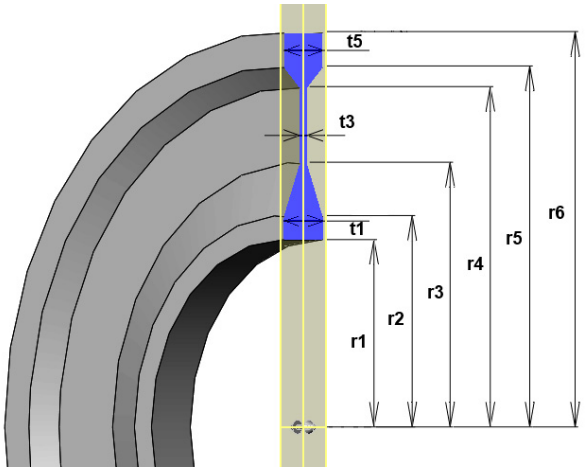

Рис. 2. Параметризация вращающегося диска

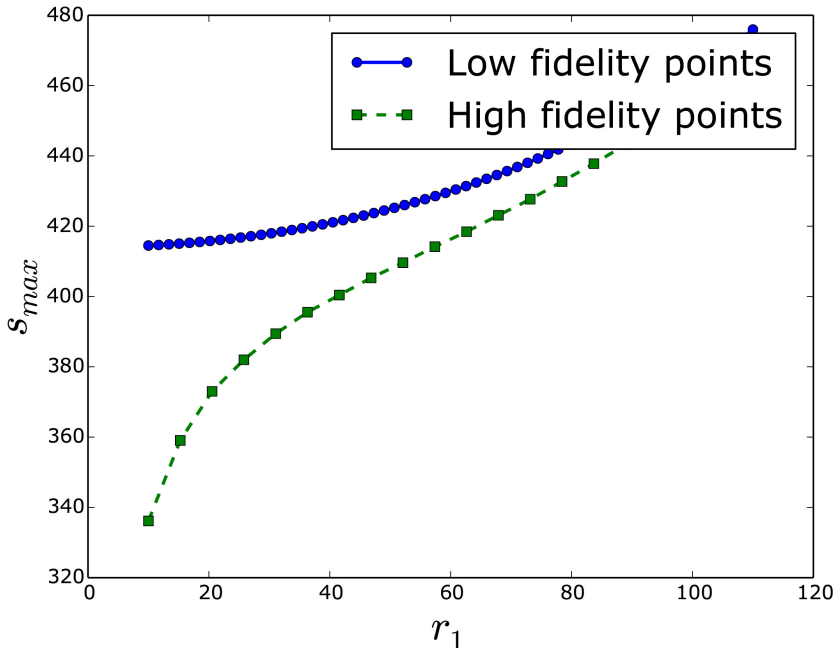

Рис. 3. Срез $s_{\max}$ по r_1 для точной (high fidelity function) и грубой (low fidelity function) функций

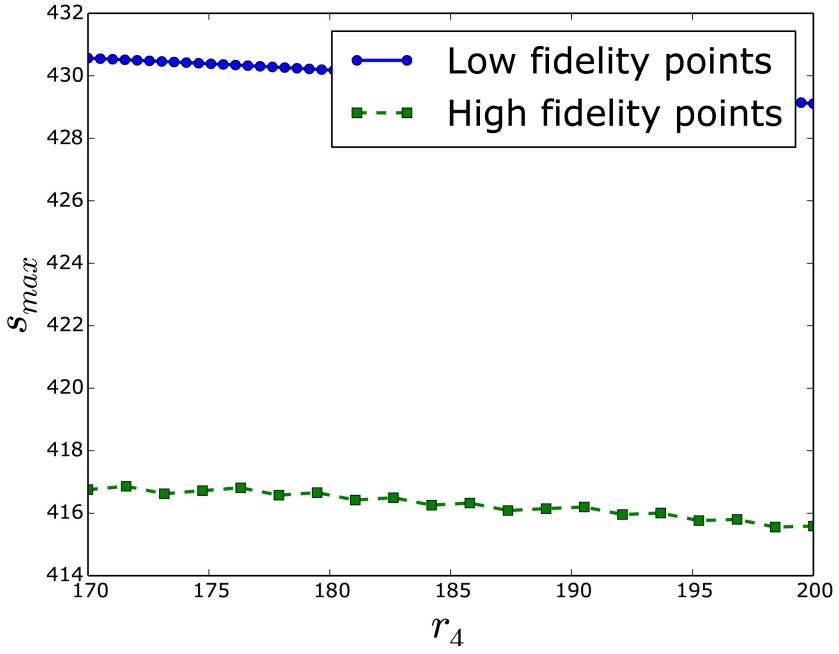

Рис. 4. Срез $s_{\max}$ по r_4 для точной (high fidelity function) и грубой (low fidelity function) функций

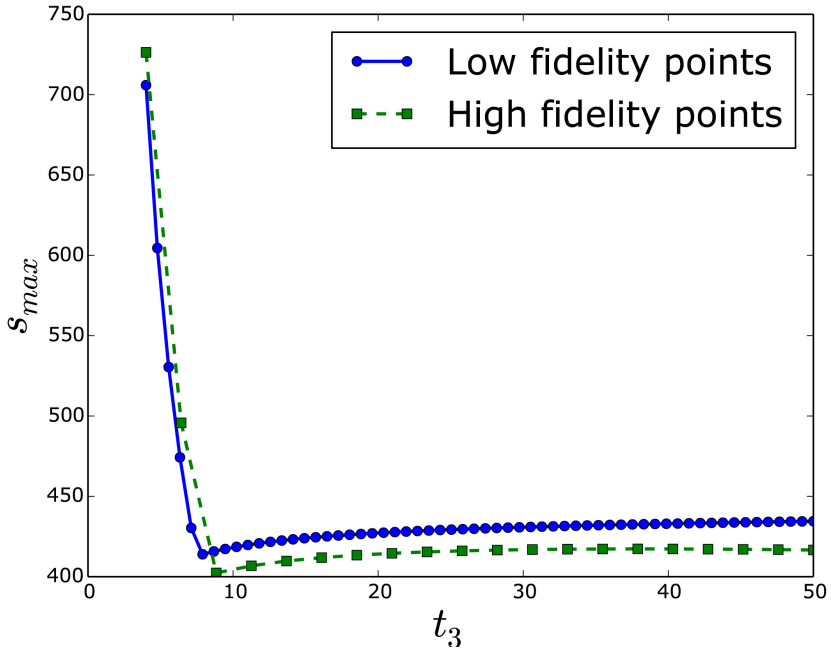

Рис. 5. Срез $s_{\max}$ по t_3 для точной (high fidelity function) и грубой (low fidelity function) функций

Сравнение качества суррогатных моделей

n_h	20	40	60	80	100
GP	0.3368	0.1826	0.1305	0.1091	0.0756
VFGP	0.1679	0.0998	0.0822	0.0564	0.0435
SVFGP	0.1018	0.0658	0.0494	0.0427	0.0339

Таблица 5. Ошибки RRMS в задаче о вращающемся диске. Выход $u_{\max}$

n_h	20	40	60	80	100
GP	0.5261	0.3181	0.2164	0.2095	0.1643
VFGP	0.2336	0.2326	0.2058	0.1321	0.1088
SFGP	0.1674	0.1095	0.1023	0.0939	0.0812

Таблица 6. Ошибки RRMS в задаче о вращающемся диске. Выход $s_{\max}$

В этом разделе мы сравниваем SVFGP с двумя базовыми методами – GP и VFGP. Для генерации точек выборки использовался метод оптимальных латинских гиперкубов. Для построения суррогатных моделей использовалось n_h подсчитанных значений точной функции, 1000 значений грубой функции для VFGP и 5000 значений грубой функции для SVFGP, из которых случайно выбиралось $n_l^1 = 1000$ базовых точек; значение n_h менялось от 20 до 100.

Для оценки качества модели использовался скользящий контроль по выборке из 140 значений точной функции (эта выборка содержала n_h точек, использованных для построения суррогатных моделей).

Результаты представлены в таблице 5 для выхода $u_{\max}$ и в таблице 6 для выхода $s_{\max}$. Видно, что SVFGP позволяет получить более точные результаты, чем VFGP, и GP, если сравнивать ошибки RRMS для этих методов.

6. ЗАКЛЮЧЕНИЕ

Мы предложили новый подход к суррогатному моделированию разноточных данных, который позволяет работать с выборками размера вплоть до нескольких десятков тысяч точек. Подход основывается на аппроксимации исходных ковариационных матриц произведениями матриц меньшего размера с помощью метода Нистрема. Получены оценка прогноза и оценка неопределенности прогноза значений точной функции. Проведено сравнение предложеного подхода с повсеместно используемыми методами на ряде реальных и модельных задач. Результаты сравнения позволяют сделать вывод о том, что с помощью предложенного подхода можно построить более точные суррогатные модели, при этом непосредственно сам процесс построения модели требует значительно меньше вычислительных затрат.

СПИСОК ЛИТЕРАТУРЫ

1. 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.
2. Noel AC Cressie and Noel A Cassie. Statistics for spatial data, volume 900. Wiley New York, 1993.
3. C. E. Rasmussen and C. K. I. Williams. Gaussian processes for machine learning. The MIT Press, 2006.
4. 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.
5. 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.

6. S Koziel, S Ogurtsov, Ivo Couckuyt, and Tom Dhaene. Cost-efficient electromagnetic-simulation-driven antenna design using co-kriging. Microwaves, Antennas & Propagation, IET, 6(14):1521–1528, 2012.

7. Edward Snelson and Zoubin Ghahramani. Sparse gaussian processes using pseudo-inputs. In Advances in Neural Information Processing Systems, pages 1257–1264, 2005.

8. Joaquin Quiñonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate gaussian process regression. The Journal of Machine Learning Research, 6:1939–1959, 2005.

9. 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.

10. Petros Drineas and Michael W Mahoney. On the nyström method for approximating a gram matrix for improved kernel-based learning. The Journal of Machine Learning Research, 6:2153–2175, 2005.

11. A.A. Zaytsev, E.V. Burnaev, and V.G. Spokoiny. Properties of the bayesian parameter estimation of a regression based on gaussian processes. Journal of Mathematical Sciences, 203(6):789–798, 2014.

12. François Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. Computational Statistics & Data Analysis, 66:55–69, 2013.

13. Gene H Golub and Charles F Van Loan. Matrix computations, volume 3. JHU Press, 2012.

14. Jeong-Soo Park. Optimal latin-hypercube designs for computer experiments. Journal of statistical planning and inference, 39(1):95–111, 1994.

15. Dervis Karaboga and Bahriye Basturk. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. Journal of global optimization, 39(3):459–471, 2007.

16. Sasan C Armand. Structural optimization methodology for rotating disks of aircraft engines. Technical report, National Aeronautics and Space Administration, Office of Management, Scientific and Technical Information Program, 1995.

17. John Charles Butcher. Numerical Methods for Ordinary Differential Equations. Wiley Online Library, 2005.

Surrogate modeling of mutlifidelity data for large samples

Burnaev E.V., Zaytsev A.A.

Typically engineers model multifidelity data (data consisting of fine and coarse functions evaluations) using Gaussian processes regression. However if the sample size exceeds few thousands points an exact inference for the Gaussian processes regression becomes computationally intractable. We propose an approximate approach based on ideas of a subset selection: we approximate covariance matrices using a subsample of the initial sample and the Nyström method. In this work we obtained formulas for a variable fidelity surrogate model prediction and a prediction uncertainty estimate along with a computation complexity of the proposed approach. Good performance for a number of real and artificial problems validates usage of the developed approach.

KEYWORDS: Multifidelity data, uncertainty estimation, Gaussian process, covariance matrix approximation, cokriging