INVARIANT APPROXIMATION AND ITS APPLICATION FOR INTEGRATION OF DATA-DOMAIN KNOWLEDGE INTO METAMODELS

A.P. Kuleshov, A.V. Bernstein, E.V. Burnaev
Kharkevich Institute for Information Transmission Problems RAS
Moscow, RUSSIA
e-mail: kuleshov@iitp.ru
Institute for Systems Analysis RAS
Moscow, RUSSIA
e-mail: a.bernstein@irias.ru
Kharkevich Institute for Information Transmission Problems RAS
Moscow, RUSSIA
e-mail: burnaev@iitp.ru

Abstract

In the present work some particular class of subject domain knowledge is formalized and the method for its integration into the process of metamodels construction is proposed. In particular, two new approaches is proposed for construction of approximation taking into account some specific invariance of the considered problem.

1 Introduction

Problem of function approximation based on data as an element of based on the data metamodel construction problem was formulated in [1]. Let M be some initial model (method) which allows for the given input $X \in \mathbf{X} \subset \mathrm{R}^p$ constructing the value $Y = F_M(X) \in \mathrm{R}^q$ of response function. Let $D_N = \{(X_i, Y_i = F_M(X_i)), i = 1, 2, \ldots, N\}$ be results of N experiments with the model M for the input data set $\mathbf{X}_N = \{X_i, i = 1, 2, \ldots, N\}$. Based on this data set an approximation function (approximator) $Y = F_{SM}(X) = F_{SM}(X|D_N)$ for the initial dependence $Y = F_M(X)$ is constructed. If for all $X \in \mathbf{X}$ (not only for $X \in \mathbf{X}_N$) an approximate equality $F_{SM}(X) \approx F_M(X)$ takes place then the new model SM defined by the constructed dependency $F_{SM}(X)$ can be considered as a surrogate for the initial model M and is named as the surrogate model (model over model, metamodel).

Quality of the constructed approximation can be estimated by the mean squared error of approximation $\varepsilon(F_{SM}|D_N) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|Y_i - F_{SM}(X_i)\|^2}$ on the learning set D_N, but at the same time constructed dependency should have "generalization ability", i.e. provide sufficient accuracy for other points $X \in \mathbf{X} \setminus \mathbf{X}_N$.

In many applications some additional a priori information about the unknown function $F_M(X)$ can also be known.

Example 1. The initial model $M = M_{stress}$ used for calculating with the help of finite element method of reserve factors Y of an aircraft stringer as a function $F_M(X)$ of geometrical shape of the stringer, material properties and forces, applied to the stringer was considered in [1].

Among the components of the p-dimensional input vector $X = (x_1, x_2, \ldots, x_p)$ there are components describing the width and length of the rectangular surface of the stringer (let say these are components x_1 and x_2 of the vector X). Then the function $F_M(X)$ has the symmetry property: $F_M(x_1, x_2, x_3, \ldots, x_p) \equiv F_M(x_2, x_1, x_3, \ldots, x_p)$.

Example 2. Let $I_1, I_2, \ldots, I_m$, $m > 1$ be some emission sources and j-th emission source is characterized by the p-dimensional parameter x_j, $j = 1, 2, \ldots, m$. Let $y_j = f(x_j), j = 1, 2, \ldots, m$ be the field created by the j-th source in the given measurement point. Measured total field $Y = F(X)$ depends on the $(p \times m)$-dimensional vector $X = (x_1, x_2, \ldots, x_m)$. If the total field Y is equal to the sum of the fields generated by each of the emission source then the function $F(X)$ has the form $F(X) = F(x_1, x_2, \ldots, x_m) = f(x_1) + f(x_2) + \ldots + f(x_m)$.

In general a priori information from the Example 1 about the approximated function $F(X)$ can be formulated as follows. Let $G = \{g\}$ be the known group consisting of finite number $|G|$ of transformations of the set of $\mathbf{X} \subset \mathrm{R}^p$ into itself. Let the function $F_M(X)$ be invariant with respect to the group G: $F(X) \equiv F(gX)$, $X \in \mathbf{X}$, $g \in G$. In Example 1 the group G of transformations of R^p into itself consists of two elements $\{g, e\}$, where $gX = (x_2, x_1, x_3, \ldots, x_p) \in \mathrm{R}^p$, and $e = g^2$ is identical transformation. It is obvious that the group G is isomorphic to the group S_2 of permutation of two elements.

Let us call the initial model for which the invariant or additive property is satisfied as the invariant (with respect to the group of transformations G) model and respectively the model with additive structure.

2 Approximation of an invariant model

Let us consider the problem of constructing the surrogate for the initial model $Y = F(X)$ having the invariant property with respect to the group of transformations G.

It is obvious that the extended sample $D_{N,G} = \{D_N(g), g \in G\}$, consisting of $N \times |G|$ elements $D_N(g) = \{(X_i, Y_i = F(gX_i)), i = 1, 2, \ldots, N\}$ can be used for constructing the approximator $F_S(X|D_{N,G})$. The mean error calculated using extended sample $D_{N,G}$ for this approximator is equal to $\varepsilon(F_S|D_{N,G}) = (\frac{1}{N \times |G|} \sum_{i=1} \sum_{g \in G} \|Y_i - F_S(gX_i)\|^2)^{1/2}$. It can be proved that

$$\varepsilon^2(F_S|D_{N,G}) = \varepsilon^2(F_{SG}|D_N) + \Delta^2(E_S, G, D_N), \tag{1}$$

here $F_{SG}(X) = \frac{1}{|G|} \sum_{g \in G} F_S(gX)$ is a "symmetrized" approximator and $\Delta^2(E_S, G, D_N) = \frac{1}{N} \sum_{i=1}^{N} V_G(X_i)$, where $V_G(X) = \frac{1}{|G|} \sum_{g \in G} (F_S(gX) - F_{SG}(X))^2 = \frac{1}{|G|} \sum_{g \in G} (F_S(gX))^2 - (F_{SG}(X))^2$. Quantities $F_{SG}(X)$ and $V_G(X)$ can be considered as a mathematical expectation $\mathrm{E}_G(F_S(gX))$ and, respectively, a variance $\mathrm{Var}_G(F_S(gX))$ of the random variable $F_S(gX)$ in which the random element g is selected in the group G uniformly randomly.

Table 1: Errors $\varepsilon\left(F_S|D_{test}\right)$ and $\varepsilon\left(F_{SG}|D_{test}\right)$ of approximators $F_S(X)$ and $F_{SG}(X)$

$\varepsilon\left(F_{S,k}|D_{test,k}\right)$	$\varepsilon\left(F_{SG,k}|D_{test,k}\right)$
4.4117E-03	3.5750E-03

Thus the error $\varepsilon\left(F_S|D_{N,G}\right)$ also can be represented as a mathematical expectation of the random error $\varepsilon\left(F_S|D_N(g)\right)$: $\varepsilon\left(F_S|D_{N,G}\right) = \mathrm{E}_G\left(\varepsilon\left(F_S|D_N(g)\right)\right)$

It follows from the equation (1) that the initial approximator $F_S(X)$ is strictly majorized by its invariant version $F_{SG}(X)$ unless only the approximator $F_S(X)$ is invariant with respect to the group G (at least on the set $\mathbf{X}_N$) when the approximators $F_S(X)$ and $F_{SG}(X)$ coincide.

Many widely used in practice procedures for construction of nonlinear approximators (Multidimensional Non-Parametric Regression, Kernel Ridge Regression, Support Vector Regression, Artificial Neural Networks, Radial Basis Function Networks etc.) construct the approximating dependencies in the form of linear combinations $\sum_{j=1}^{p} \alpha_j \times h(X, \beta_j)$ of non-linear functions $h(X, \beta)$ from the chosen "dictionary". For such approximators the invariant approximator also has the form of linear combination $\sum_{j=1}^{p} \alpha_j \times H_G(X, \beta_j)$ of functions from the dictionary, consisting of the symmetrized functions $H_G(X, \beta) = \mathrm{E}_G\left(h(gX, \beta)\right)$.

3 Approximation of an additive model

Let the vector $X = (x_1 : x_2 : \ldots : x_m) \in \mathrm{R}^{pm}$ be represented as m sequentially written subvectors $\{x_1, x_2, \ldots, x_m\}$ with the same dimension p. From the additive property it follows that for the vector $x \in \mathrm{R}^p$ $f(x) = \frac{1}{m}F(x : x : \ldots : x)$, for any vector $x \in \mathrm{R}^p$ and consequently $F(X) = F(x_1 : x_2 : \ldots : x_m) = \frac{1}{m}\sum_{i=1}^{m} F(x_i : x_i : \ldots : x_i)$.

It is obvious that the models with additive structure are invariant with respect to the group of transformations G_m of R^{pm} into itself which is isomorphic to the permutation group S_m with the order m: if $g = (i_1, i_2, \ldots, i_m)$ is some permutation of the numbers $(1, 2, \ldots, m)$ then $gX = (x_{i_1} : x_{i_2} : \ldots : x_{i_m})$.

Let $F_S(X)$ be some initial approximator and $F_{SL}(X) = F_{SL}(x_1 : x_2 : \ldots : x_m) = \frac{1}{m}\sum_{i=1}^{m} F_S(x_i : x_i : \ldots : x_i)$ be its "additivited" version. It is obvious that the approximator $F_{SL}(X)$ is invariant with respect to the group G_m.

If the approximator $F_S(X)$ has the form $\sum_{j=1}^{p} \alpha_j \times h(X, \beta_j) \equiv \sum_{j=1}^{p} \alpha_j \times h((x_1 : x_2 : \ldots : x_m), \beta_j)$, then the approximator $F_{SL}(X)$ has the form $\sum_{j=1}^{p} \alpha_j \times H_L(X, \beta_j)$, where $H_L(X, \beta) = \frac{1}{m}\sum_{i=1}^{m} h((x_i : x_i : \ldots : x_i), \beta)$.

4 Results of computational experiments

Efficiency of proposed procedures was analyzed using computational experiments.

In Example 1 learning sample D_{learn} consisting of $N_{learn} = 50000$ results of experiments with model M_{stress} was obtained. Using this sample both the approximator

Table 2: Errors $\varepsilon(F_{S,k}|D_{test,k})$, $\varepsilon(F_{SG,k}|D_{test,k})$ and $\varepsilon(F_{SL,k}|D_{test,k})$, $k = 1, 2, 3, 4$ of approximators $F_{S,k}(X)$, $F_{SG,k}(X)$ and $F_{SL,k}(X)$, $k = 1, 2, 3, 4$

k	$\varepsilon(F_{S,k}\|D_{test,k})$	$\varepsilon(F_{SG,k}\|D_{test,k})$	$\varepsilon(F_{SL,k}\|D_{test,k})$
1	0.0263	0.0099	0.0037
2	0.0421	0.0171	0.0052
3	0.0446	0.0209	0.0072
4	0.0510	0.0294	0.0098

$F_S(X)$ proposed in [2] and its invariant version $F_{SG}(X)$ were constructed. Mean errors $\varepsilon(F_S|D_{test})$ and $\varepsilon(F_{SG}|D_{test})$ of these approximators $F_S(X)$ and $F_{SG}(X)$ were calculated for new testing sample D_{test} consisting of $N_{test} = 100000$ results of other independent experiments. The values of these errors are listed in the Table 1.

Let us consider for Example 2 several training samples $D_{train,k}$, $k = 1, 2, 3, 4$ consisting of the same number $N_{train} = 10000$ of measuring results of total field generated by $m = 2$ emission sources. In the k-th series of measurements the sets $\mathbf{X}_{learn,1}$ and $\mathbf{X}_{learn,2}$ consisting of the $p = 7$ dimensional values of characteristics x_1 and x_2 of the emission sources I_1 and I_2 were chosen randomly in the set $\mathbf{X}_k \subset \mathrm{R}^p$, $k = 1, 2, 3, 4$. These sets differ in a spread of the emission sources positions and in their frequency characteristics. Using generated learning sets both the approximator $F_{S,k}(X)$ proposed in [2] and its invariant $F_{SG,k}(X)$ and additive versions $F_{SL,k}(X)$, $k = 1, 2, 3, 4$ versions were constructed, $k = 1, 2, 3, 4$. Mean errors $\varepsilon(F_{S,k}|D_{test,k})$, $\varepsilon(F_{SG,k}|D_{test,k})$ and $\varepsilon(F_{SL,k}|D_{test,k})$ of the approximators $F_{S,k}(X)$, $F_{SG,k}(X)$ and $F_{SL,k}(X)$ were calculated for new testing samples $D_{test,k}$, $k = 1, 2, 3, 4$ each consisting of $N_{test} = 50000$ results of other independent experiments in which the characteristics of emission sources were also chosen randomly in the sets $\mathbf{X}_k \subset \mathrm{R}^p$. The normalized (with respect to the output's range) values of the errors are listed in the Table 2.

Obtained results clearly demonstrate an efficiency of use of data-domain knowledge in surrogate models.

References

[1] Bernstein A.V., Kuleshov A.P. (2010). Computer Data Analysis in Metamodeling. *In the present book.*

[2] Burnaev E.V., Grihon S. (2009). Construction of the Metamodeles in Support of Stiffened Panel Optimization. *Extended abstracts of VI International Conference "Mathematical Methods in Reliability. Theory. Methods. Applications" (MMR-2009)*, June 22-29, Moscow, Russia. pp. 124 - 128.

[3] Burnaev E.V., Belyaev M.G., Prihodko P.V. (2010). Approximation of multidimensional dependency based on an expansion in parametric functions from the dictionary. *In the present book.*