Creation of delay-operators for multiscale forecasting by means of symbolic regression

Kulunchakov A. S.

Moscow Institute of Physics and Technology

Draft of introduction

Suppose that one needs to build a forecasting machine for a response variable. Given a large set of time series, one can advance a hypothesis that they are related to this variable. Relying upon this hypothesis, we can use given time series as features for the forecasting machine. However, the values of time series could be produced with different frequencies. Therefore, we should take into account not only the values, but the delays as well.

The simplest model for forecast is a linear one. In the presence of large set of features this model can approximate the response quite well. To avoid the problem of multiscaling, we introduce a definition of delay-operators. Each delay-operator corresponds to one time series and represents continuous correlation function. This correlation function shows a dependence between the response variable and corresponding time series. Therefore, each delay-operator put weights on the values of corresponding time series depending on the greatness of the delay.

Having these delay-operators, we avoid the problem of multiscaling. To find them, we use genetic programming and symbolic regression.

If the resulted weighted linear regression model would produce poor approximation, we can use a nonlinear one instead. To find good nonlinear function, we would use symbolic regression as well.

Statement

Denote a large set of time series $\mathcal{D} = \{\boldsymbol{s}^i\}_{i=1}^{i=N}$, where N is a number of time series. The values of a response variable are denoted as $\boldsymbol{s}^0$ Each time series $\boldsymbol{s}^k$ (including the response variable $\boldsymbol{s}^0$) is a set of observations $\boldsymbol{s}^k = \{(t_j^k, s_j^k)\}_{j=1}^{j=N(k)}$ with its own number of observations $N(k)$, where s_j^k-s represent values and t_j^k-s represent timestamps for these values.

Given a primitive set $\mathcal{G} = \{f_1, \ldots, f_m, x_1\}$ consisting of mathematical functions (they may be parametric), we can construct superpositions of them. The space of such superpositions is denoted by $\mathcal{F}$. Each delay-operator $d_k(\boldsymbol{w}, x_1)$ lies in $\mathcal{F}$ and represents valid mathematical formula. As each delay-operator $d_k(\boldsymbol{w}, x_1)$ should represent a good approximation for the correlation function between $\boldsymbol{s}_k$ and $\boldsymbol{s}_0$, the parameters of $d_k(\boldsymbol{w}, x_1)$ are tuned to maximize

$$\sum_{i=1}^{N(0)/2} \left(s_i^0 - \bar{\boldsymbol{s}}^0\right)\left(s_i^{k0} - \bar{\boldsymbol{s}}^{k0}\right),$$

where $\boldsymbol{s}^{k0}$ is a new time series with the same timestamps with $\boldsymbol{s}^0$ and values calculated as follows

$$s_j^{k0} = d_k(\boldsymbol{w}, t_j^0 - t_m^k) \cdot s_m^k,$$

where m is an index of the last available value from $\boldsymbol{s}^k$ up to the timestamp t_j^0 of the response variable. To comment on more precisely, we take the closest previous value from $\boldsymbol{s}_k$, calculate its delay $t_j^0 - t_m^k$ from considered timestamp t_j^0 of the response variable s^0 and finally weight it by $d_k(\boldsymbol{w}, t_j^0 - t_m^k)$.

Once the parameters are tuned, we can use $d_k(\boldsymbol{w}, x_1)$ to weight delayed values of s^k. So, we can build a linear model to forecast values of $\boldsymbol{s}^0$.

$$\boldsymbol{s}^0 = \sum_{i=1}^{N} p_i \cdot \boldsymbol{s}^{i0} + \varepsilon,$$

where each new time series $\boldsymbol{s}^{i0}$ consists of values of $\boldsymbol{s}^i$ weighted according to their delays:

$$s_j^i = d_k(\boldsymbol{w}, t_j^0 - t_m^k) \cdot s_m^k, \;\; m = \mathop{\arg\max}_{t_j^0 - t_\pi^k \geqslant 0}(\pi)$$

Note, that we tune parameters on the first half of s^0. Therefore, the prediction error is measured on the second half

$$\text{MSE} = \sum_{i=N(0)/2}^{N(0)} \left(s_i^0 - \hat{s}_i^0\right)^2.$$

To find explicit expressions of delay-operators we use genetic programming and symbolic regression. Note that moreover, we can also apply them to find a function f, which describe the dependence between $\boldsymbol{s}^0$ and $\boldsymbol{s}^k$ well

$$\boldsymbol{s}^0 = f(\boldsymbol{s}^{10}, \dots, \boldsymbol{s}^{i0}, \dots, \boldsymbol{s}^{N0}) + \varepsilon,$$

References