ISSN 1407-7493 COMPUTER SCIENCE DATORZINĀTNE 2009-7493

APPLIED COMPUTER SYSTEMS
LIETIŠĶĀS DATORSISTĒMAS

A COMPARISON OF SUBSET SELECTION AND ADAPTIVE BASIS FUNCTION CONSTRUCTION FOR POLYNOMIAL REGRESSION MODEL BUILDING

APAKŠKOPAS ATLASES UN ADAPTĪVĀS BĀZES FUNKCIJU KONSTRUĒŠANAS SALĪDZINĀJUMS POLINOMU REGRESIJAS MODEĻU BŪVĒŠANĀ

Gints Jēkabsons, Dr.sc.ing., researcher, Riga Technical University, Meza 1/3, LV-1048, Riga, Latvia, gintsj@cs.rtu.lv

Jurijs Lavendels, Dr.sc.ing., professor, Riga Technical University, Meza 1/3, LV-1048, Riga, Latvia, jurisl@cs.rtu.lv

Polynomial regression, regression modelling, basis function construction, model ensembling, heuristic search

1. Introduction

In regression modelling commonly polynomial models are used. Polynomials are very flexible and often used when there is no theoretical model available. To obtain a polynomial model that describes the relations in data sufficiently well and does not overfit, some kind of model building must be performed. Typically it is done using the subset selection approach (also called variable selection) [1] where the goal is from a fixed full set of predefined basis functions to find the best subset that gives the best predictive performance of a respective regression model. Usually, the full set of basis functions is equal to the set of basis functions in a "full" polynomial model of a user-predefined complexity (i.e. degree).

To find the best subset, it must be searched for. Searching through all the possible subsets needs exponential runtime and thus is impractical in most cases. Hence in subset selection heuristic search methods are used. They efficiently traverse the space of subsets, by adding and deleting basis functions and use an evaluation measure that directs the search into areas of increased performance. The typical examples of the search methods are Forward Selection (also known as Sequential Forward Selection, SFS) and Backward Elimination (also known as Sequential Backward Selection, SBS) [1,2,3].

The approach of subset selection assumes that the chosen fixed full set of predefined basis functions contains a subset that is sufficient to describe the target relation sufficiently well. However, in most cases the necessary full set is not known and needs to be guessed (e.g. by specifying the degree of the "full" polynomial model) since it will differ from one data set to another. In many cases (especially when the studied dependencies are very complex and not well studied) that means either a non-trivial (and long) trial and error process or acceptance of a possibly inadequate model.

In [4,5] we considered an approach for polynomial model building that is different from the subset selection – letting the regression model building method itself construct the basis functions necessary for creating the model without restricting oneself to the basis functions of a predefined full model. This is achieved by replacing the standard refinement operators of subset selection, namely the addition and deletion of the basis functions, with other operators that allow not only adding or deleting the basis functions but also allow modifying them. In this manner all the needed basis functions are adaptively constructed during the heuristic search process trading-off the simplicity and predictive performance of the models. Hence we call this approach Adaptive Basis Function Construction (ABFC) [5]. The approach allows generating polynomials of arbitrary complexity, does not require the

user to predefine any basis functions (or to set a maximal degree) for model generation, and, in addition, allows using most of the same heuristic search algorithms and evaluation measures which are used in subset selection methods. In [4,5,6] we proposed also two instances of the approach.
In the present paper we compare the two approaches for polynomial regression model building – subset selection and ABFC – both theoretically and empirically in terms of their underlying principles, computational complexity, and predictive performance. Additionally in empirical evaluations we compare the two instances of ABFC also to two other well-known regression modelling methods.
Sections 2 and 3 shortly review the subset selection approach and the ABFC approach. Section 4 deals with the theoretical comparisons of the two approaches. Section 5 briefly reviews the two instances of the ABFC. Section 6 deals with the empirical comparisons.

2. Subset selection

A regression model describes a relation between a vector of d real-valued input variables $x = (x_1, x_2, ..., x_d)$ and a single real valued output variable y. Using a finite number n of training observations $(x_{(i)}, y_{(i)})$, $i = 1,2,...,n$ one wants to build a model F that allows predicting the output value for yet unseen input values as closely as possible [1].
Generally, a polynomial regression model may be defined as a linear summation of basis functions:

$$F(x) = \sum_{i=1}^{k} a_i f_i(x) \tag{1}$$

where a are model's parameters; k is the number of basis functions (equal to the number of model's parameters); $f_i(x)$ ($i = 1,2,...,k$) is a polynomial basis function. The estimation of model's parameters is made based on the training data typically using the ordinary least-squares method [1] minimizing

$$a = \arg\min_{a} \sum_{i=1}^{n} \left(y_{(i)} - F(x_{(i)})\right)^2 . \tag{2}$$

In the subset selection approach if the full set of predefined basis functions is ϕ then any of its subsets $f \subseteq \phi$ can be specified as a binary vector s with cardinality $|s| = m$ (i.e. the total number of predefined basis functions) in which the j-th value equal to 1 means that the j-th basis function ϕ_j is included in the subset forming the model but a value equal to 0 means that it is not:

$$f = \{\phi_j \mid j = 1,2,...,m;\ s_j = 1\}. \tag{3}$$

Usually, the full set of basis functions ϕ is equal to the set of basis functions in a "full" polynomial model of a user-predefined (maximal) degree p (a non-negative integer). Hence, for example, if $d = 3$ and $p = 2$, then the full set contains $m = 10$ basis functions

$$\phi = \{1, x_1, x_2, x_3, x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, x_2x_3\}$$

and the vector

$$s = (1,0,1,0\,1,1,0,0\,1,0)^T$$

corresponds to the (ordered) subset

$$f = \{1, x_2, x_1^2, x_2^2, x_1x_3\}$$

which in turn corresponds to the model

$$F(x) = a_1 + a_2x_2 + a_3x_1^2 + a_4x_2^2 + a_5x_1x_3 .$$

Formally the problem of finding the best subset f^* can be defined as finding the best combination of values for *m* bits in the vector *s*:

$$s^* = \arg\min_s J\left(\{\phi_j \mid j = 1,2,...m;\ s_j = 1\}\right) \tag{4}$$

where $J(.)$ is a criterion which evaluates the predictive performance of the regression model resulting from the subset.

Searching through all the possible combinations of *m* bits requires evaluation of 2^m models, i.e. exponential runtime and thus is impractical in most cases. Hence in subset selection heuristic search methods are used. They efficiently traverse the space of subsets, by adding and deleting basis functions and use an evaluation measure that directs the search into areas of increased performance.

Considering [7,8,9], in order to characterize a heuristic search problem one must define the following: 1) initial state of the search; 2) available state-transition operators; 3) search strategy; 4) evaluation measure; 5) termination condition.

In the subset selection approach for polynomial regression, typically the *initial states* are models that correspond to the empty subset, the subset with only the intercept term in it, full subset of all the defined basis functions, or a randomly chosen subset; the typical *state-transition operators* are addition and deletion of a basis function; the typical *search strategy* is the hill climbing [2,3,7] which in combination with the empty subset initial state and the addition operator becomes SFS but in combination with the full subset initial state and the deletion operator becomes SBS; the classical *evaluation measures* are the statistical significance tests [1], however, currently two other strategies predominate: employment of complexity penalization criteria (e.g. the Akaike's Information Criterion, AIC [10]) and the resampling techniques (e.g. Hold-Out, Cross-Validation (CV), and Bootstrap [11, 1]); the *termination condition* typically corresponds to finding of a state in which none of the state-transition operators can lead to a better state (i.e. a local minima).

However, before the actual search for the best subset is performed the user must predefine all the basis functions in the full set ϕ (or set a sufficiently large maximal degree *p*) which can turn into a non-trivial (and long) trial and error process (see section 5).

3. Adaptive basis function construction

The basis functions in a polynomial regression model generally can be defined as a product of original input variables each raised to some degree:

$$f_i(x) = \prod_{j=1}^{d} x_j^{r_{ij}} \tag{5}$$

where *r* is a $k \times d$ matrix of non-negative integer degrees such that r_{ij} is the degree of the *j*-th variable in the *i*-th basis function. Note that when all r_j's of a basis function are equal to 0, we have the intercept term.

Given a number of input variables *d*, matrix *r* with a specified number of rows *k* and with specified values of each of its element completely defines a model with all its basis functions. The set of basis functions included in a model is then equal to

$$f = \left\{ \prod_{j=1}^{d} x_j^{r_{i,j}} \mid i = 1,2,...,k \right\}. \tag{6}$$

For example, if $d = 3$ and $k = 4$, then the matrix

$$r = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

corresponds to the set

$$f = \left\{x_1^0 x_2^0 x_3^0,\ x_1^1 x_2^0 x_3^0,\ x_1^0 x_2^1 x_3^3,\ x_1^1 x_2^1 x_3^1\right\} = \left\{1,\ x_1,\ x_2 x_3^3,\ x_1 x_2 x_3\right\}$$

which in turn corresponds to the model

$$F(x) = a_1 x_1^0 x_2^0 x_3^0 + a_2 x_1^1 x_2^0 x_3^0 + a_3 x_1^0 x_2^1 x_3^3 + a_4 x_1^1 x_2^1 x_3^1 = \\ = a_1 + a_2 x_1 + a_3 x_2 x_3^3 + a_4 x_1 x_2 x_3.$$

Formally the problem of finding the best subset f^* can be defined as finding the best combination of non-negative integer values of elements in the matrix *r*:

$$r^* = \arg\min_r J\left(\left\{\prod_{j=1}^{d} x_j^{r_{i,j}} \mid i = 1,2,...,k\right\}\right). \tag{7}$$

As neither the upper bounds of *r* elements' values nor the upper bound of *k* are defined, it is possible to generate polynomial models of arbitrary complexity, i.e. of arbitrary number of basis functions each with arbitrary degree for each input variable.

Finding the best combination of values in *r* requires searching. In the context of a heuristic search problem, we now define the five (see Section 2) components for the ABFC approach.

The *initial states* can be models that correspond to the empty subset or to the subset with only the intercept term in it. However, choosing initial state as the full set or as a randomly generated subset is not directly applicable as in ABFC the state space is infinite.

Using efficient *state-transition operators* is vital for the search process to be successful. The main methodological difference between the subset selection approach and the ABFC approach is exactly in the state-transition operators used. Generally there are two different basic types of modifications to an existing polynomial model: adding/deleting the basis functions (as in subset selection approach) and operating with the degrees of variables in an existing basis function (e.g. increasing or decreasing them). In contrast to the subset selection approach, in the ABFC both types of modifications are used, replacing the two standard state-transition operators of subset selection with five operators allowing generation of any polynomial which can be defined using the matrix *r*.

Here are brief descriptions of the five state-transition operators of ABFC. *Operator1*: Addition of a new basis function with one of the degrees set to one (and all the others set to zero). *Operator2*: Increasing of one of the degrees in one of the existing basis functions by one. *Operator3*: Addition of an exact copy of an already existing basis function with one of the degrees increased by one. *Operator4*: Decreasing of one of the degrees in one of the existing basis functions by one. *Operator5*: Deleting of one of the existing basis functions. For more information on the state-transition operators of ABFC, see [4,5].

We categorize the listed state-transition operators of ABFC as complication operators (the first three) and simplification operators (the last two). If the search is started from an empty or some small set of functions, the complication operators do the main job – they "grow" the model. The simplification operators on the other hand work as purifiers – they decrease the unnecessarily high degrees and delete the unnecessary basis functions.

In relation to *search strategies* most of the heuristic algorithms that are applicable to subset selection can also be used in ABFC. This is achieved by treating the complication and simplification operators

as the addition and deletion operators (correspondingly) of the subset selection approach (see [4,5] for more details).
As *evaluation measures* the same predominating ones of subset selection can be efficiently used also in ABFC: complexity penalization criteria and resampling techniques. The former in contrast to the latter usually does not require high computational resources, allows one to use all the available data for training, as well as is less noisy (creating less local minima in the state space). One of the issues on the inapplicability of statistical significance tests in ABFC is the requirement of the tests for all the compared models to be nested. However, because of the *Operator2* and *Operator4* the models are not always nested. For more details on evaluation measures in ABFC consult [4,5].
The typical *termination condition*, that is met when the search locates a state in that none of the refinement operators can lead to a better state, is a natural choice also in the ABFC.
A more detailed description of the approach is given in our previous studies [4,5].

4. Comparison of subset selection and ABFC

The approach of subset selection assumes that the chosen fixed full set of predefined basis functions contains a subset that is sufficient to describe the target relation sufficiently well. However, in most cases the necessary full set (or the maximal degree p) is not known and needs to be guessed since it will differ from one data set to another. In many cases that means either a non-trivial (and long) trial and error process restarting the search many times, each time considering a different value of p, or acceptance of a possibly inadequate model.
Moreover, the choice of p for practical reasons is also guided by the amount of time needed for the search procedure. Using one or both of the two basic state-transition operators, already in the very first iteration of the search the order of the branching factor of the current state in the state space is equal to the number of basis functions in the full set: $O(m) = O(d^p)$ [1]. This means that for a fixed number of input variables increase of the maximal degree exponentially increases the number of models to be evaluated in each iteration of the search. For larger values of d and p this can render almost any kind of subset selection impractical or at least make it so slow that there are only limited possibilities finding the right value for p even with a very simple and fast search algorithm. This in turn means that in practical applications to perform a search for a good model in acceptable time one needs to consider lower values of p when there is increase of d. However, there is no reason to believe that generally the nonlinearity of y will indeed decrease when d is increased.
In ABFC the branching factor of a state in the state space depends on the number of input variables d and on the number of basis functions already in the model k. The upper bound of the number of possible modifications to a model for *Operator1* is equal to d; for *Operator2*, *Operator3*, and *Operator4* it is equal to dk; for *Operator5* it is equal to k. So the upper bound on the branching factor is of order $O(d + 3dk + k) = O(dk)$ which is linear in respect to both d and k. Beginning the search with a model with only the intercept term included, the branching factor of the current state is equal d, increasing in each next iteration together with k.
Assuming that the "best" model found by the search includes a total of k_* basis functions while in each iteration the number of basis functions in the current model was increased by 1, it is concluded that in subset selection using one or both basic state-transition operators the number of evaluated states is of order

$$O\left(\sum_{i=1}^{k_*} d^p\right) = O(d^p k_*). \tag{8}$$

However, in ABFC the number of evaluated states is of order

$$O\left(\sum_{i=1}^{k_*} di\right) = O\left(dk_* \sum_{i=1}^{k_*} i\right) = O\left(dk_* \frac{k_*(k_*+1)}{2}\right) = O\left(dk_*^3 + dk_*^2\right) = O(dk_*^3). \tag{9}$$

Hence it is expected that in the context of required computational resources for model building the ABFC approach is (asymptotically) more efficient than the subset selection approach when the following inequality is satisfied:

$$d^p k_* > dk_*^3. \tag{10}$$

Overall, one can conclude that, in comparison to the subset selection approach, the efficiency of the ABFC approach increases together with the number of input variables and the required nonlinearity of the model (the value of p) and decreases together with the complexity k_* of the "best" found model. Moreover the relative efficiency of ABFC is multiplied when in subset selection the required value of p is unknown and needs to be guessed.
The characteristic properties of both approaches including those discussed in Section 2, Section 3 and the current section are summarized in Table 1.

Table 1

A summarization of characteristic properties of the subset selection and ABFC approaches

Property	**Subset selection approach**	**ABFC approach**	**Notes on the properties of the ABFC**
Basis functions for model building	A predefined full set of basis functions or a chosen maximal degree p (both need to be guessed or searched for multiplying the required computational resources)	The basis functions are adaptively constructed during the model building process	There is no need to predefine the full set of basis functions or the maximal degree (and the model building process is done only once)
The fundamental principle of the model building process	A search for the best subset of a predefined full set of basis functions	The basis functions are adaptively constructed during the model building process	The needed basis functions are obtained as a part of the solution; increased elasticity is obtained
Limitation of the model's degree and/or definition of basis functions for inclusion	Mandatory	An additional option	
A possible way of defining a specific state in the state space (model)	A binary-valued vector of length m	A $k \times d$ matrix of non-negative integers	
Initial state	Any state (empty set, full set, randomly chosen set)	The simplest model in the space or some other sufficiently simple model	By setting constraints it is possible also to start from any other state
State-transition operators	Typically, addition and deletion	Complication and simplification	Apart from the operators given in Section 3, it is possible to develop also many other kinds

Search strategy	All the search algorithms typical for subset selection [1,2,3]	Almost all the same algorithms are adaptable also for ABFC [4,5]	
Evaluation measure	Statistical significance tests, complexity penalization criteria, resampling methods	Complexity penalization criteria, resampling methods	
Typical termination condition	Finding of a state in which none of the state-transition operators can lead to a better state	Finding of a state in which none of the state-transition operators can lead to a better state	
Number of states in the state space (i.e. the number of all possible combinations of basis functions in the model)	$O(2^m)$ or $O(2^{d^p})$	Infinite	If required, by setting corresponding constraints it is possible to make the state space finite
Branching factor of a state in a state space	$O(d^p)$ – exponential	$O(dk)$ – linear	The relative efficiency of the ABFC approach (asymptotically) increases together with the number of input variables and the required nonlinearity of the model
The number of evaluated states if the complexity of the "best" found model is k_*	$O(d^p k_*)$ – exponential	$O(dk_*^3)$ – polynomial	The relative efficiency of the ABFC approach (asymptotically) decreases together with the complexity of the "best" found model

5. Instances of the ABFC: F-ABFC and EF-ABFC

In [4,5] we proposed an instance of the ABFC approach – a regression model building method called Floating Adaptive Basis Function Construction (F-ABFC). In the context of heuristic search, the method has the following properties: *initial state* corresponds to the subset with only the intercept term in it; the used *state-transition operators* are all the five listed operators of the ABFC; as the *search strategy* the Sequential Floating Forward Selection [12] is used is (hence the name of the model building method); as the *evaluation measure* the Corrected Akaike's Information Criterion (AICC) [10] is used; the *termination condition* corresponds to finding of a state in which none of the state-transition operators can lead to a better state. A more detailed description of the F-ABFC is given in [4,5].

As discussed in [6], there are two issues that plague the methods of model building in polynomial regression, especially when working with relatively small data samples: the selection bias and the selection instability. In practice, both these issues are usually ignored frequently resulting in models of lower predictive performance.

Selection bias occurs when in the search procedure one uses the same data to compute values of models' parameters and also to evaluate the models for selection purposes [11,13,14,15]. This also includes usage of resampling techniques – because the search procedure is evaluating so many subsets,

it is likely that some of them lead to models that have high accuracy for the validation set but low accuracy for the test set [11]. In any case, the more intensive the search procedure, the larger the selection bias will be leading to overfitted models.
The other issue, selection instability (also called selection variance), is related to the fact that small perturbations of the data can lead to vastly different subsets of the basis functions (frequently corresponding to either underfitted or overfitted models), because of getting stuck in different local minima or because of noisy model evaluation methods [16,17,18].
In [6] we proposed a model building method called Ensemble of Floating Adaptive Basis Function Construction (EF-ABFC). In EF-ABFC we tried to deal with these issues in the F-ABFC by using a collaboration of two techniques: 1) CV performed over the entire F-ABFC search process for selection of one best model from the best models of each F-ABFC iteration (the validation set is not used for model evaluation during the search, instead it is used strictly only for post-evaluation and post-selection after the search process has already ended). This post-evaluation can detect whether the search process at some iteration might have started to generate overfitted models and select a model of some earlier iteration that is hopefully not (or at least less) overfitted; 2) the v models from the v CV loops are combined using a simple ensemble technique – unweighted model averaging [19]. Combining models in this way can have the effect of smoothing out erratic models that overfit the data and gain more stability in the modelling process [13,17,19].
A practical application of the EF-ABFC can be found in e.g. [10].

6. Empirical experiments

The main goal of the performed experiments was to compare the both instances of ABFC to a widely used instance of subset selection SFS used together with the AICC criterion, as well as to full polynomials (FP) and to two other well known regression modelling methods – Multivariate Adaptive Regression Splines (MARS) [21] and Locally Weighted Polynomials (LWP) [20,22]. We compared the methods in terms of both, predictive performance of the induced regression models as well as necessary computational resources. The performance of the methods is evaluated on 11 different regression data sets from the UCI repository (http://archive.ics.uci.edu/ml/), namely, autoMPG, autoPrice, bodyfat, fishcatch, friedman, housing, housingNOX, machineCPU, pbc, servo, stock. They are chosen because of the relatively low number of data cases, which is also common in real practical situations, as well as because of mostly continuous input variables.
All the experiments were performed on Pentium 4 2.4GHz computer with Hyper Threading turned on. Note that the time consumptions presented in the table are rather rough measurements as the methods are implemented with different levels of optimization of calculations. In the experiments we used our in-house software with implementations of EF-ABFC, F-ABFC, SFS, LWP, and full polynomials. We used piecewise-cubic MARS version 3.6 without a specific restriction of the number of basis functions or interaction orders and with the number of degrees of freedom for the models' parameters selected using 10-fold CV. For LWP we used the Gaussian weight function setting the bandwidth parameter and the polynomial degree using Leave-One-Out CV.
Since for the subset selection method the required degree p is not known before the search for the best model is performed, we used three types of degree selection: 1) always use a fixed degree (titled "SFS + AICC, p = …" where "p = …" denotes the used degree); 2) start from the first degree and increase it as long as the value of AICC improves (titled "SFS + AICC + AICC" as both the search and choice of the degree were performed using AICC); 3) start from the first degree and increase it as long as the value of 10-fold CV improves (titled "SFS + AICC + CV" as the search is still performed using AICC but the degree is chosen using CV).
In the experiments we estimated predictive performance of the built models on unseen data samples using 10-fold CV and averaged the results. Note that for all the methods this CV was done as an outer loop over the entire model building process and, in each CV iteration, the set aside test set was used strictly only for the final evaluation of the built model. Note also that here it is important to distinguish

between the CV inside the EF-ABFC, MARS, LWP, or SFS+AICC+CV and the CV used for evaluation of model building methods – the CV loops are completely separated.
The predictive performance of a model in test data set is measured in terms of Relative Root Mean Squared Error (the lower the value of RRMSE the more accurate the model):

$$RRMSE = 100\% \sqrt{\sum_i \left(y_{(i)} - F(x_{(i)})\right)^2} \Big/ \sqrt{\sum_i \left(y_{(i)} - \bar{y}\right)^2} \tag{11}$$

where $\bar{y}$ is the mean of the observed values. Note that prior to dividing the data sets into CV folds, the order of the cases was randomized.
Table 2 presents the results of the performed experiments. Note that the maximal degree of the full polynomials here is only 2 as for some data sets higher degrees were impossible because of too low number of data cases. However, the polynomials obtained by the SFS were up to the sixth degree (results for $p = 5$ and $p = 6$ are not shown as they were not built for all data sets because the predictive performance started to deteriorate).

Table 2

The results of the performed empirical experiments averaged over all 11 data sets

Method	Average RRMSE	Average elapsed time
Full polynomial, $p = 1$	50.97	-
Full polynomial, $p = 2$	$9 \cdot 10^{15}$	-
SFS + AICC, $p = 1$	50.96	< 0.1 s
SFS + AICC, $p = 2$	41.23	4.5 s
SFS + AICC, $p = 3$	39.54	3.6 min
SFS + AICC, $p = 4$	48.10	36 min
SFS + AICC + AICC	50.82	46 min
SFS + AICC + CV	36.54	5.3 h
F-ABFC	44.36	2.0 min
EF-ABFC	34.14	13 min
MARS	41.29	4.1 min
LWP	36.91	5.3 min

The results confirm the superiority of EF-ABFC over F-ABFC as well as over the results obtained by subset selection in terms of predictive performance. Additionally, while EF-ABFC here is about 6.5 times slower than F-ABFC, it is still about 24.5 times faster than the best subset selection method. Comparing EF-ABFC to MARS and LWP, the models of EF-ABFC are of higher predictive performance while the time consumption is also slightly higher. Overall, one can conclude that the model building methods of ABFC outperform the methods of subset selection at least considering the required computational resources and compare rather well to all the other employed methods.

7. Conclusion

In the paper we compared two approaches for polynomial regression model building – subset selection and ABFC – both theoretically and empirically in terms of their underlying principles, computational complexity, and predictive performance. In the subset selection approach before the actual search for the best subset is performed the user must predefine all the basis functions in the full set or specify a sufficiently large maximal degree. It was concluded that the task can turn into a non-trivial and computationally complex trial and error process which was confirmed in the empirical experiments. EF-ABFC, which gave the best predictive performance, was about 24.5 times faster than the best considered subset selection method. Overall, it was concluded that the model building methods of ABFC outperform the methods of subset selection at least considering the required computational

resources and compare rather well to all the other employed methods. They can be viewed as competitive tools for regression modelling.

References

1. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. – Berlin: Springer, 2003.
2. Webb A.R., Statistical Pattern Recognition, 2nd ed. – John Wiley & Sons, 2002. – P.496.
3. Dash M., Liu H. Feature Selection for Classification // Intelligent Data Analysis. An International Journal. – Elsevier, Vol. 1, 1997. – P.131-156.
4. Jekabsons G., Lavendels J. An approach for polynomial regression modelling using construction of basis functions // Scientific Proceedings of Riga Technical University, 5, Computer Science, 34. – Riga: RTU 2008.
5. Jekabsons G., Lavendels J. Polynomial regression modelling using adaptive construction of basis functions // IADIS International Conference Applied Computing 2008. – Algarve, Portugal, 2008. – P.269-276.
6. Jekabsons G. Ensembling adaptively constructed polynomial regression models // International Journal of Intelligent Systems and Technologies (IJIST), Vol. 3, No 2. – WASET, 2008. – P.56-61. (http://www.waset.org/ijist/v3/v3-2-11.pdf)
7. Russell S.J., Norvig P. Artificial intelligence: a modern approach, 2nd ed. – Englewood Cliffs, New Jersey: Prentice Hall, 2002.
8. Molina L.C., Belanche L., Nebot A. Feature selection algorithms: a survey and experimental evaluation // Proceedings of the International Conference on Data Mining (ICDM'02). – Maebashi City: IEEE Computer Society, 2002. – P.306-313.
9. Ginsberg M.L. Essentials of artificial intelligence. – Morgan Kaufmann, 1993.
10. Hurvich C.M., Tsai C.-L. Regression and time series model selection in small samples // Biometrika, Vol. 76. – 1989. – P.297-307.
11. Kohavi R., John G.H., Wrappers for feature subset selection // Artificial Intelligence, vol. 97. – 1997. – P.273-324.
12. Pudil P., Ferri F.J., Novovicova J., Kittler J. Floating search methods for feature selection with nonmonotonic criterion functions // Proceedings of the International Conference on Pattern Recognition, Vol. 2. – Los Alamitos, CA: IEEE, 1994. – P.279-283.
13. Elder IV J.F. The generalization paradox of ensembles // Journal of Computational and Graphical Statistics, Vol. 12. – 2003. – P.853-864.
14. Reunanen J. Overfitting in making comparisons between variable selection methods // Journal of Machine Learning Research, Vol. 3. – 2003. – P.371-382.
15. Loughrey J., Cunningham P. Overfitting in wrapper-based feature subset selection: the harder you try the worse it gets // 24rth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI-2004). – 2004. – P.33-43.
16. Harrell Jr. F.E. Regression modelling strategies with applications to linear models, logistic regression, and survival analysis. – New York: Springer, 2001.
17. Breiman L. Heuristics of instability and stabilization in model selection // Annals of Statistics, Vol. 24. – 1996. – P.2350-2383.
18. Kotsiantis S., Pintelas P. Combining Bagging and Boosting // International Journal of Computational Intelligence, Vol. 1. – 2004. – P.324-333.
19. Opitz D., Maclin R. Popular Ensemble Methods: An Empirical Study // Journal of Artificial Intelligence Research, Vol. 11. – 1999. – P.169-198.
20. Kalnins K., Ozolins O., Jekabsons G. Metamodels in design of GFRP composite stiffened deck structure // Proceedings of 7th ASMO-UK/ISSMO International Conference on Engineering Design Optimization, Association for Structural and Multidisciplinary Optimization in the UK (ASMO-UK). – Bath, UK, 2008. – P.11. (in print)
21. Friedman J.H. Fast MARS, Department of Statistics, Stanford University, Tech. Report LCS110, 1993.

22. Cleveland W., Loader C. Smoothing by Local Regression: Principles and Methods (with discussion) // Computational Statistics. – 1995.

Jēkabsons G., Lavendels J. Apakškopas atlases un adaptīvās bāzes funkciju konstruēšanas salīdzinājums polinomu regresijas modeļu būvēšanā

Polinomu regresijas modeļu būvēšanā apakškopas atlases pieeja pieņem, ka izvēlētā fiksētā pilnā iepriekš definēto bāzes funkciju kopa ir pietiekama, lai pietiekami labi aprakstītu datos esošās sakarības. Tomēr vairumā gadījumu nepieciešamā bāzes funkciju kopa nav zināma un to ir jāuzmin, kas var nozīmēt netriviālu (un ilgu) mēģinājumu un kļūdu procesu. Mūsu iepriekšējā pētījumā tika apskatīta polinomu regresijas modeļu būvēšanas pieeja, kas atšķiras no apakškopas atlases – tā ļauj regresijas modeļu būvēšanas metodei pašai veidot bāzes funkcijas, kas nepieciešamas, lai izveidotu jebkuras sarežģītības modeli, neaprobežojoties ar iepriekš definēta pilna modeļa bāzes funkcijām. Šī pieeja ir nosaukta par Adaptīvu bāzes funkciju konstruēšanu (Adaptive Basis Function Construction, ABFC). Šajā rakstā mēs salīdzinām abas minētās polinomu regresijas modeļu būvēšanas pieejas – apakškopas atlase un ABFC – gan teorētiski, gan empīriski, to pamatprincipu, skaitļošanas sarežģītības un prognozēšanas spējas nozīmē. Empīriskajos eksperimentos ABFC ir salīdzinātas arī ar divām citām labi zināmām regresijas modelēšanas metodēm – lokāli svērtajiem polinomiem un daudzfaktoru adaptīvajiem regresijas splainiem.

Jekabsons G., Lavendels J. A comparison of subset selection and adaptive basis function construction for polynomial regression model building

The approach of subset selection in polynomial regression model building assumes that the chosen fixed full set of predefined basis functions contains a subset that is sufficient to describe the target relation sufficiently well. However, in most cases the necessary set of basis functions is not known and needs to be guessed – a potentially non-trivial (and long) trial and error process. In our previous research we considered an approach for polynomial regression model building which is different from the subset selection – letting the regression model building method itself construct the basis functions necessary for creating a model of arbitrary complexity without restricting oneself to the basis functions of a predefined full model. The approach is titled Adaptive Basis Function Construction (ABFC). In the present paper we compare the two approaches for polynomial regression model building – subset selection and ABFC – both theoretically and empirically in terms of their underlying principles, computational complexity, and predictive performance. Additionally in empirical evaluations the ABFC is compared also to two other well-known regression modelling methods – Locally Weighted Polynomials and Multivariate Adaptive Regression Splines.

Екабсон Г., Лавендел Ю. Сравнение отбора подмножества и адаптивного строительства базисных функция в постройке модели полиномной регрессии

Подход отбора подмножества в постройке модели полиномной регрессии предполагает, что выбранный фиксированный полный набор предопределенных базисных функций содержит подмножество, которое является достаточным для описания целевой связи достаточно точно. Тем не менее в большинстве случаев необходимый набор базисных функций не известен и должен быть угадан – потенциально нетривиальный (и долгий) процесс проб и ошибок. В наших предыдущих исследованиях мы рассматривали подход постройки модели полиномной регрессии, который отличается от выбора подмножества – в этом подходе методу постройки модели регрессии позволено самому строить базисные функций, необходимые для создания модели произвольной сложности, не ограничивая себя базисными функциями заранее предопределенной полной модели. Этот подход назван Адаптивним строительством базисных функций (Adaptive Basis Function Construction, ABFC). В данной статье мы сравниваем эти два подхода постройки модели регрессии – отбор подмножества и ABFC – как теоретически, так и эмпирически с точки зрения их основных принципов, вычислительной сложности и способности прогнозировния. Кроме того в эмпирических экспериментах ABFC сравнивается также с двумя другими хорошо известными методами регрессионного моделирования – локально взвешиванные полиномы и многомерные адаптивные регрессивые сплайны.