POLYNOMIAL REGRESSION MODELLING USING ADAPTIVE CONSTRUCTION OF BASIS FUNCTIONS

Gints Jekabsons, Jurijs Lavendels
Institute of Applied Computer Systems, Riga Technical University
Meza 1/3, LV-1048, Riga, Latvia

ABSTRACT

The approach of subset selection in regression modelling 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 the paper we consider an adaptive basis function construction approach that in many problems has a potential to be more efficient. It lets the modelling method itself construct the basis functions necessary for creating a regression model of arbitrary complexity with adequate predictive performance. We also introduce an instance of the approach that as a search strategy uses the floating search algorithm. To evaluate the proposed method, we compare it to other regression modelling methods, including the well-known Sequential Forward Selection, on artificial and real world data.

KEYWORDS

Polynomial regression, subset selection, basis function construction, heuristic search.

1. INTRODUCTION

Various applications in engineering, statistics, computer science, health sciences, and social sciences are concerned with estimating "good" predictive models from available data. In such problems the goal is to estimate unknown dependency (or model) from training data, in order to use this model for predicting future samples. A model describes the relation between a multidimensional input, x, and an output, y. If the y is a continuous variable, the task is called a regression task and the model is called a regression model.

In regression commonly polynomial models are used. Polynomials are very flexible and often used when there is no theoretical model available. Such model may be defined by a linear summation of basis functions:

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

where a_i are model's parameters; k is the number of the *used* basis functions (equal to the number of model's parameters); $f_i(x)$ are the basis functions that generally may be defined as a product of original input variables each raised to some order:

$$f_i(x) = \prod_{j=1}^{d} x_j^{r_{ij}} \tag{2}$$

where d is the number of the original input variables; r_{ij} is the order of the j-th variable in the i-th basis function (a non-negative integer). Note that when all r_j's of a basis function are equal to 0, we have the intercept term. The estimation of parameters a_i of the regression models is made based on a finite number, n, of training data cases, $(x_{(1)}, y_{(1)})$, $(x_{(2)}, y_{(2)})$, …, $(x_{(n)}, y_{(n)})$, typically using the ordinary least-squares method, OLS.

The regression model to use should be neither too simple (too low number of basis functions), causing underfitting, nor too complex (too high number of basis functions), causing overfitting. Otherwise model's ability to generalize to new data will be relatively poor.

To obtain a polynomial regression model, that describes the relations in data sufficiently well, typically the subset selection approach (also called variable selection) (Rawlings, 1998) is used where the goal is from the fixed full set of predefined basis functions to find the best subset that gives the best predictive performance of the regression model.

Usually, in the subset selection approach the full set of basis functions is chosen such that the order of each possible polynomial model does not exceed a previously chosen highest allowed order, p, i.e., each $r_{ij} \in \{0,1,...,p\}$ and $\sum_{j=1}^{d} r_{ij} \leq p$ for all i. Then the number of *all* defined basis functions is

$$m = \prod_{i=1}^{p} (1 + d / i) \tag{3}$$

and the number of all possible subsets, from which we want to find the best, is equal to 2^m.

In order to find the best subset some kind of search must be performed. Searching through all possible subsets needs exponential runtime and thus is impractical in most cases. Hence 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 the Forward Selection (also known as Sequential Forward Selection, SFS) and the Backward Elimination (also known as Sequential Backward Selection, SBS) (Rawlings, 1998; Aha & Bankert, 1996).

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 set of basis functions is not known and needs to be guessed (e.g., by specifying the maximal order, p, of the resulting polynomial) since it will differ from one data set to another. In many cases that means a non-trivial (and long) trial and error process, in that one can generate sets of functions, working with which, in some problems of even moderate dimensionality, may become computationally too demanding even for the heuristic search methods (as we already demonstrated in (Jekabsons & Lavendels, 2007) and will demonstrate in the empirical experiments of Section 4). A more convenient and potentially also a more efficient way would be to let the modelling method itself construct the basis functions necessary for creating the regression model with adequate predictive performance.

In this paper, we consider an adaptive polynomial regression modelling approach with automatic construction of basis functions using heuristic search – Adaptive Basis Function Construction (ABFC). The approach allows generating polynomials of arbitrary complexity and does not require the user to predefine any basis functions for model creation. We also introduce an instance of the approach – a new regression model induction method – Floating ABFC (F-ABFC).

Note that the ABFC approach is partially a generalization of the ideas from a rather recently published work (Todorovski et al., 2004) where the authors introduce to a polynomial equation induction method called Constrained Induction of Polynomial Equations for Regression (CIPER) that was developed in a context of differential equation discovery, inductive databases, and constraint-based data mining. CIPER can also be viewed as an instance of the ABFC approach. However, it has some drawbacks regarding too high sensitivity to local minima and the nesting effect (Jekabsons & Lavendels, 2007), that we tried to reduce in F-ABFC.

To evaluate the introduced polynomial regression model induction method, we empirically compared it to different other regression model induction methods using three artificial problem data sets, three real-world data sets, and one metamodelling data set.

The rest of this paper is organized as follows. In Section 2 we describe the general framework of the ABFC approach. This is done mainly from the viewpoint of heuristic search. In Section 3 we shortly characterize the F-ABFC. An empirical comparison of the proposed method with the other methods is presented in Section 4. Finally Section 5 concludes the work and gives directions of possible future research.

2. A GENERAL FRAMEWORK FOR THE ADAPTIVE BASIS FUNCTION CONSTRUCTION APPROACH

In the ABFC approach, the standard refinement operators of subset selection, namely the addition and deletion of the basis functions, are replaced with other operators that not only allow adding or deleting the basis functions but also allow changing the basis functions themselves. In Figure 1, there is shown relation between subset selection and ABFC. Subset selection operates with a string of bits of constant length (column named “Included”) where each bit indicates whether a predefined basis function (columns “Function” and “Form”) is present (“1”) or absent (“0”). ABFC approach on the other hand operates directly with the orders of each variable in each function as well as creates new functions as necessary (column “Matrix of orders, r”). Thus, in ABFC the search operates directly with the dynamically-sized matrix r in the

equation (2). This results in an infinite space of candidate regression models and we can generate polynomials of arbitrary complexity without predefining any basis functions.

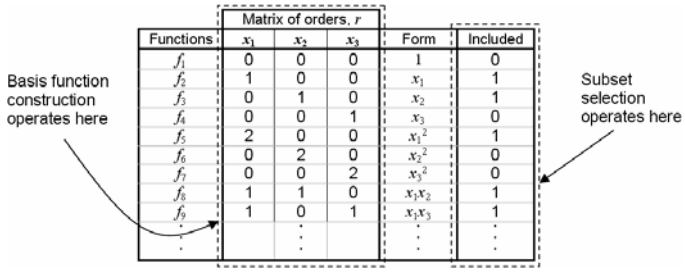

Figure 1. Relation between subset selection and ABFC

Summarizing (Ginsberg, 1993; Blum & Langley, 1997; Russell & Norvig, 2002), in order to characterize the 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. Note that in the rest of this paper instead of the term "state-transition operator" we will use the term "(model) refinement operator" (as in (Todorovski et al., 2004)), that is somewhat more convenient in the context of regression modelling.

In the context of polynomial regression subset selection, 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 available *refinement operators* are addition and deletion of any one basis function; the *search strategies* are the successive addition of basis function (as in SFS) or successive removal of them (as in SBS); the *evaluation measure*s are either the statistical significance tests (Rawlings, 1998), the complexity penalization criteria (e.g., the Akaike's Information Criterion (Akaike, 1974, Hurvich & Tsai, 1989; Burnham & Anderson, 2002)), or the resampling techniques (e.g., Hold-Out or Cross-Validation (Kohavi, 1995; Rawlings, 1998)); the *termination condition* typically corresponds to finding of state in that none of the refinement operators can lead to a better state.

A more detailed comparison of subset selection and ABFC is given in our previous study (Jekabsons & Lavendels, 2007). In the next subsections the five characteristics of the heuristic search process in the context of ABFC will be described.

2.1 Initial State of the Search and the Termination Condition

As the state space has become infinite, a natural initial state of the search is now the state where the subset of basis functions is empty. Or one may also choose a state that has some very few simple functions (e.g., just one function that corresponds to the intercept term or all the functions that correspond to the first order terms). A small set of simple functions might be also generated randomly. In our research so far we only considered the initial state to be the subset with the intercept term (moreover, this function stays in the model at all times and is not allowed to be modified or deleted).

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 of course a natural choice also in the ABFC.

2.2 Refinement Operators

Using efficient refinement operators is vital for the search process for the best regression model to be successful. In this section we discuss some of the possible refinement operators.

Generally there are two different basic ways to refine an existing model: adding/deleting the basis functions and operating with the orders of variables in an existing basis function (e.g., increasing or decreasing them). Here are the five considered refinement operators. *Operator1*: Addition of a new basis function with one of the orders set to one (and all the others set to zero). *Operator2*: Increasing of one of the orders in one of the existing basis functions by one. *Operator3*: Addition of an exact copy of already existing basis function with one of the orders increased by one. *Operator4*: Decreasing of one of the orders in one of the existing basis functions by one. *Operator5*: Deleting of one of the existing basis functions.

Additionally for each refinement operator, except Operator5, special care is taken to prevent basis function duplicates in the resulting model.

Note that if the one function that corresponds to the intercept term is allowed not to be in the model, the Operator1 must be extended to be able to also generate the intercept term (i.e., addition of a new basis function with all the orders set to zero). Note also that if it is allowed to use this basis function for construction of other basis functions, the Operator1 in certain situations will generate the same basis functions as Operator3 (i.e., when the current subset has the intercept term included), becoming obsolete. However, to maintain generality, instead of abandoning the Operator1 we simply do not allow the intercept term to be used for construction of other functions.

We categorize the listed refinement operators 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 regression model. The simplification operators on the other hand work as purifiers – they decrease the unnecessarily high orders and delete the unnecessary basis functions.

The first two complication operators were already introduced in (Todorovski et al., 2004) where they were used in CIPER. However, using only these two operators three issues can arise, all of which can lead to getting stuck in local minima too early, that in turn can result in poor predictive performance of the resulting regression model: 1) In the first iterations the branching factor of the state space may be too low – the search algorithm may have too few choices to be able to continue the search. 2) Operator2 increases an order in an existing function but does not take into consideration the possibility that both versions of the basis function may be needed. The only way to reconstruct the lost function is to start from scratch – by using the Operator1. 3) Without the use of simplification operators a regression model may contain unnecessarily high orders and include unnecessary basis functions that may prevent truly necessary modifications (the so-called nesting effect (Pudil et al., 1994)).

The first two issues are addressed by the Operator3 as it does both increases the branching factor and allows copying the basis functions. The last one issue is addressed by the two simplification operators.

2.3 Search Strategy

Most search strategies that are applicable to subset selection also can be used in ABFC. One exception is the strategies that start their search from the full subset (e.g., SBS). As in ABFC approach there exists no full subset, the only way to make these strategies work would be to define the full subset to be some sufficiently large set of functions, definition of which mostly would have no special reason, and moreover, it possibly would bring us back to the approach of subset selection. Another exception is the strategies that require the state-representing data structures to be of constant length and are not generally biased towards simpler models (e.g., the chromosomes in most Genetic Algorithms). However with appropriate modifications they might become applicable.

In this paper, we will consider only the directly applicable search strategies the simplest of which is the SFS. However, SFS moves only forward, in direction of more complex models, so it would use only the complication operators. Hence the Steepest Descent Hill Climbing (Russell & Norvig, 2002), Plus-a Take Away-b (Stearns, 1976), and Sequential Floating Forward Selection (Pudil et al., 1994) strategies may be considered, all of which allow also the backward moves (see Section 3 where we introduce the F-ABFC with search strategy based on SFFS).

2.4 Evaluation Measure

The evaluation of regression models corresponding to alternative subsets of basis functions is classically done by statistical significance testing (Rawlings, 1998). However, currently two other model selection strategies predominate: employment of complexity penalization criteria (e.g., the Akaike's Information Criterion (Akaike, 1974; Hurvich & Tsai, 1989; Burnham & Anderson, 2002)) and resampling techniques (e.g., Hold-Out or Cross-Validation (Kohavi, 1995; Rawlings, 1998)). The complexity penalization criteria in contrast to the resampling techniques usually do not require high computational resources and allow one to use all the available data for training. The most widely known complexity penalization criterion is Akaike's Information Criterion (AIC) (Akaike, 1974; Burnham & Anderson, 2002). However in the current research

we use its small sample corrected version (AICC) (Hurvich & Tsai, 1989; Burnham & Anderson, 2002) defined as follows:

$$AICC = n\log(MSE) + 2k + (2k(k+1))/(n-k-1), \quad (4)$$

where MSE is Mean Squared Error in training data. For problems with small *n* AICC is suited better than AIC but converges to AIC as *n* becomes large (Hurvich & Tsai, 1989; Burnham & Anderson, 2002). Note that the best fitting model is that whose AICC value is the lowest.

Because of the fact that during the search the orders of the variables in the basis functions may be increased without changing the actual number of basis functions (using the Operator2), AICC may tolerate unnecessary increase of orders in basis functions, as it cannot detect a change in model's structure if the number of parameters stayed the same. To deal with this problem, while preserving AICC's original form, we used additional penalization using Akaike's weights (Akaike, 1974; Burnham & Anderson, 2002) in the very moment of comparison.

When using a refinement operator that increases the order of a variable without changing the number of model's basis functions, we ask the Akaike's weight w_{new} of the newly constructed model to be at least 10% higher than that of the "old" model. In other words, the weight (that also may be directly interpreted as the conditional probability of the new model being better than the competitor) of the new model should be >60% instead of >50%:

$$w_{new} = \frac{\exp(-0.5\Delta AICC)}{1+\exp(-0.5\Delta AICC)} > 0.6 \quad (5)$$

where $\Delta AICC = AICC_{new} - AICC_{old}$.

3. AN INSTANCE OF THE APPROACH: F-ABFC

In this section, we shortly characterize a new method for polynomial regression model induction that is an instance of the ABFC approach – F-ABFC. It uses all the five refinement operators listed in Section 2.2 and, to maintain simplicity and efficiency, as the search strategy that of the floating search method SFFS (that already showed its advantages in (Pudil et al., 1994; Jekabsons & Lavendels, 2007)) is used – hence the name of the F-ABFC.

Here is an overview of F-ABFC's characteristics: *Initial state*: the subset with one function that corresponds to the intercept term (this function stays in the model at all times and is not allowed to be modified or deleted). *Refinement operators*: all five listed in Section 2.2. *Search strategy*: based on SFFS. *Evaluation measure*: AICC (see Section 2.4). *Termination condition*: when no further improvements are possible. A more detailed description of the F-ABFC is given in our previous study (Jekabsons & Lavendels, 2007).

Additionally, if there indeed is a sufficient reason to believe that the regularities in data will be captured best with a model order of which does not exceed a certain value, of course the ABFC approach too allows limiting the order of the models. This can be done for example in the following two ways: 1) by setting the maximal order of the basis functions just like in the subset selection approach, i.e., $\sum_{j=1}^{d} r_{ij} \le p$ for all *i*; 2) by setting the maximal order of each particular input variable in the basis functions, i.e., $r_{ij} \le p_j$ for all *i*, where p_j is maximal order of the *j*-th input variable. However, note that in the empirical experiments described in this paper we did not use any kind of limiting.

4. EMPIRICAL EXPERIMENTS

The main goal of the performed experiments was to compare the F-ABFC with the popular subset selection methods SFS and SFFS used together with AICC, another instance of the ABFC – CIPER, as well as all possible simple full polynomials starting from the second order as long as the number of parameters is not higher than the number of training data cases. 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 three artificial problem data sets, three data sets from the WEKA

project website (http://www.cs.waikato.ac.nz/ml/weka/), and one data set from an application in metamodelling of glass fibre bar stability described in (Jekabsons et al., 2007).

All the experiments were performed on Pentium 4 2.4GHz computer with Hyper Threading turned on. Note that the time consumption presented in the tables is only a rough measurement as the methods are implemented in different software and with different levels of optimization of calculations. In the experiments we used our in-house software with implementations of SFS, SFFS, F-ABFC, and CIPER. In all these methods the AICC was used. This allowed us to compare the use of refinement operators and search strategies without any hindrance because of the different criteria. As an implementation of the original CIPER (implemented with a different criterion (Todorovski et al., 2004)) we used the original software that is publicly available at http://ai.ijs.si/pljubic/ciper/ciper.html, kindly provided by the authors of the method. In both versions of CIPER we used the default beam width, which is 16.

In experiments with the artificial problem data sets, we randomly generated training data set with 150 cases with all the input variables uniformly distributed over the interval [0,1], and tested the induced regression models on unseen test data set of 10000 randomly generated cases. We repeated the process for 10 different training data sets and averaged the results. In all the other experiments we estimated predictive error of the induced models on unseen data samples using 10-fold Cross Validation (CV) and averaged the results. The predictive performance of a model in test data set is measured in terms of Relative Root Mean Squared Error, RRMSE:

$$RRMSE = RMSE/STD = \sqrt{\sum_i \left(y_{(i)} - \hat{y}_{(i)}\right)^2} \Big/ \sqrt{\sum_i \left(y_{(i)} - \bar{y}_{(i)}\right)^2} \tag{6}$$

where $\hat{y}_{(i)}$ is the corresponding predicted value for the observed value $y_{(i)}$; $\bar{y}$ is the mean of the observed values. While RMSE (Root Mean Square Error) represents model's deviation from the data, the STD (Standard Deviation) captures how irregular the problem is. The lower the value of RRMSE, the more accurate the model.

The three artificial problem data sets were generated using the following functions:

$$F_1 = \prod_{i=1}^{7} x_i + \prod_{i=4}^{10} x_i + \varepsilon$$

$$F_2 = \prod_{i=1}^{7} x_i + \prod_{i=4}^{10} x_i + x_{10} \prod_{i=4}^{10} x_i + \varepsilon$$

$$F_3 = \sin(\pi x_1 / 2)\sin(x_2)\cos(x_3)\cos(x_4) + \sin(\pi x_5 / 2)\sin(x_6)\cos(x_7)\cos(x_8) + \varepsilon$$

where ε is a normal noise with mean equal to zero and standard deviation $\sigma = 0.001$ for the first two functions and $\sigma = 0.01$ for F_3. If the function that generated the data is assumed to be unknown, regression model induction by subset selection approach with the F_1 and F_2 data sets might be considered impractical. With F_1 high predictive performance can be reached only by defining the maximal order of polynomials to be at least $p = 7$ (this of course would need to be guessed or obtained by experimenting) creating $m = 43758$ basis functions (according to equation (3)) that leads us to a state space with number of states equal to $2^{43758} \approx 3 \cdot 10^{13172}$. With F_2 it would be $2^{109395} \approx 1.5 \cdot 10^{32931}$ as the maximal order must be at least $p = 8$.

Table 1 presents the results of experiments with the regression model induction methods used on the artificial problem data sets. Here the methods of ABFC are clearly superior – they were able to relatively quickly find models with high predictive performance. With F_2, both versions of CIPER did not perform so well – by using only the first two refinement operators they could not construct the third basis function that is an exact copy of the second basis function having the order of x_{10} increased by one. As the third basis function would need to be constructed from scratch, the problem is that, even if the method would not stuck in local minima, at some iteration there may exist two exact copies of one basis function, which is of course not allowed. The F-ABFC, on the other hand, manages very well – after creation of the second basis function it just uses the *Operator3* to create a copy of it with the order of the x_{10} increased by one.

It should be noted that with the first two artificial problems the models found by F-ABFC contain slightly more basis functions than necessary. This is explainable by the fact that the AICC criterion does not assume the true model to be among the candidates and may give preference to more complex models (Burnham & Anderson, 2002; Hurvich & Tsai, 1989).

The three data sets from the WEKA project website are the following: "autoMpg" (392 cases, 7 input variables), "bodyfat" (252 cases, 14 input variables), and "housing" (506 cases, 13 input variables). 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. The data set from the metamodelling problem had

300 cases and 4 input variables. For all these data sets before dividing them into CV folds, the order of the cases was randomized.

The results on the WEKA data sets, presented in the Table 2, confirm that the ABFC methods can have approximately the same or better predictive performance as the subset selection methods, without the necessity to choose the maximal order. They also show that, in common with the subset selection methods, when the goal quantity of the basis functions in the model is relatively low (as with the first two WEKA data sets), the needed computational resources are much lower.

With the "housing" data set, CIPER gets stuck in local minima. However, the nesting effect of CIPER is most notable with the first two WEKA data sets – the number of basis functions in the models induced by CIPER is approximately equal to those induced by F-ABFC but the predictive performance is lower.

Comparing the CIPER+AICC and the original CIPER, it can be observed that the criterion of the latter forces to construct much simpler models, making the method much faster and more robust to overfitting, but on the other hand making the method to underfit (as can be slightly noticed again with the first two data sets).

Table 1. Results for the artificial problems. Average RRMSE error (%), elapsed time (s), and *k*

	F_1		F_2		F_3	
Method	**RRMSE**	**Time / *k***	**RRMSE**	**Time / *k***	**RRMSE**	**Time / *k***
Full, $p = 2$	67.33	- / 66	66.09	- / 66	15.02	- / 45
SFS, $p = 2$	71.82	1.8 / 28	74.27	0.8 / 18	14.84	0.8 / 29
SFS, $p = 3$	64.60	39 / 47	65.66	39 / 48	8.62	13 / 42
SFS, $p = 4$	54.55	320 / 62	58.61	419 / 69	8.31	70 / 47
SFS, $p = 5$	40.42	2251 / 78	43.93	2074 / 77	7.07	212 / 47
SFFS, $p = 2$	73.05	1.4 / 19	74.50	1.1 / 15	14.90	1.5 / 29
SFFS, $p = 3$	66.03	35 / 36	68.42	29 / 34	8.47	19 / 39
SFFS, $p = 4$	57.64	416 / 58	58.56	391 / 55	6.93	51 / 35
SFFS, $p = 5$	41.41	2842 / 75	42.43	1831 / 64	8.05	245 / 42
CIPER original	0.75	1.5 / 4	11.63	2.9 / 5	11.44	12 / 15
CIPER+AICC	3.94	80 / 17	9.69	150 / 20	11.57	186 / 27
F-ABFC	2.33	1.3 / 9	1.95	1.4 / 9	6.59	34 / 32

Table 2. Results for the WEKA and the metamodelling data sets. Average RRMSE error (%), elapsed time (s), and *k*

	autoMpg		**bodyfat**		**housing**			**Metamodelling**	
Method	**RRMSE**	**Time / *k***	**RRMSE**	**Time / *k***	**RRMSE**	**Time / *k***	**Method**	**RRMSE**	**Time / *k***
Full, $p = 2$	37.17	- / 36	23.05	- / 120	41.14	- / 105	Full, $p = 2$	88.44	- / 15
Full, $p = 3$	54.29	- / 120	-	- / -	-	- / -	Full, $p = 3$	72.09	- / 35
Full, $p = 4$	~75000	- / 330	-	- / -	-	- / -	Full, $p = 4$	48.02	- / 70
Full, $p = 5$	-	- / -	-	- / -	-	- / -	Full, $p = 5$	25.97	- / 126
Full, $p = 6$	-	- / -	-	- / -	-	- / -	Full, $p = 6$	19.33	- / 210
SFS, $p = 2$	36.77	0.2 / 11	15.30	1.1 / 13	41.72	29 / 48	SFS, $p = 2$	87.88	< 0.1 / 10
SFS, $p = 3$	37.28	3.1 / 18	14.14	11 / 13	38.52	997 / 83	SFS, $p = 3$	76.33	0.4 / 19
SFS, $p = 4$	37.82	28 / 26	14.87	157 / 16	53.63	7600 / 97	SFS, $p = 4$	46.45	3.8 / 34
SFS, $p = 5$	36.21	59 / 20	37.72	1902 / 17	-	- / -	SFS, $p = 5$	30.75	22 / 49
SFFS, $p = 2$	36.56	0.3 / 9	16.10	1.4 / 11	42.16	31 / 32	SFS, $p = 6$	16.12	105 / 70
SFFS, $p = 3$	36.84	2.7 / 14	14.14	11 / 11	41.88	615 / 54	SFS, $p = 7$	11.69	404 / 86
SFFS, $p = 4$	37.93	17 / 16	15.85	147 / 12	39.02	3636 / 57	SFS, $p = 8$	9.02	981 / 102
SFFS, $p = 5$	36.57	41 / 16	14.81	1727 / 12	-	- / -	SFS, $p = 9$	11.72	1963 / 113
CIPER original	38.31	1.2 / 7	14.71	1.2 / 6	45.21	23 / 14	CIPER original	66.32	4.7 / 21
CIPER+AICC	38.03	37 / 14	14.46	58 / 12	46.34	2850 / 33	CIPER+AICC	60.47	60 / 23
F-ABFC	37.06	2.9 / 13	12.81	2.4 / 11	36.91	1059 / 53	F-ABFC	3.46	2405 / 99

With the last data set from the metamodelling problem, it was observed that here the highest predictive performance is the property of the complex high-order polynomials. The F-ABFC here is clearly superior to all the others – it was able to find models with the best predictive performance.

Overall, the results confirm that model induction using the ABFC has a potential to adaptively generate polynomial regression models of adequate predictive performance without the requirement to predefine any basis functions as it is in the subset selection approach. It also can be observed that the F-ABFC shows the most superiority when the resulting regression model needs to be highly nonlinear.

5. CONCLUSION

In this paper, we considered an adaptive basis function construction approach in polynomial regression modelling. The approach is different from the standard subset selection approach in that it does not require any basis functions to be predefined. Instead the needed basis functions are adaptively constructed using heuristic search. We also introduced an instance of the approach – a new regression model induction method. The performed empirical experiments showed that the method has a potential to efficiently construct regression models of relatively high predictive performance. We showed also that it can be efficiently applied in metamodelling problems outperforming the typical methods of this application – full polynomials and SFS.

Directions of future research include developing further generalizations of the refinement operators to gain more control over the thoroughness of the search process. This generalization could involve increasing/decreasing orders not just by one but by some dynamically changing small number, value of which could be set in reaction to some events in the current iteration of the search (e.g., significant changes in the number of found models that are better than the current model). This way it could be possible to lower the probability of getting stuck in local minima or make the search faster.

ACKNOWLEDGEMENT

This work has been partly supported by the European Social Fund within the National Program "Support for the carrying out doctoral study program's and post-doctoral researchers" project "Support for the development of doctoral studies at Riga Technical University".

REFERENCES

Aha, D.W., Bankert, R.L., 1996. A comparative evaluation of sequential feature selection algorithms. Learning from Data, Fisher, D., Lenz, H.J. (eds.), Springer, New York, USA, pp. 199-206.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, pp. 716-723.

Blum, A.L., Langley, P., 1997. Selection of relevant features and examples in machine learning. Artificial Intelligence, 97, pp. 245-271.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer

Ginsberg, M.L., 1993. Essentials of Artificial Intelligence. Morgan Kaufmann

Hurvich, C.M., Tsai, C-L., 1989. Regression and time series model selection in small samples. Biometrika, 76, pp. 297-307.

Jekabsons, G. and Lavendels, J., 2007. An approach for polynomial regression modelling using construction of basis functions. Scientific Proceedings of Riga Technical University, Computer Science. RTU, Riga, Latvia (in print)

Jekabsons, G. et al., 2007. Polynomials in metamodeling of glass fiber bar stability. Scientific Proceedings of Riga Technical University, Computer Science. RTU, Riga, Latvia (in print)

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, pp. 1137-1145.

Pudil, P. et al., 1994. Floating search methods in feature selection. Pattern Recognition Letters, 15, pp. 1119-1125.

Rawlings, J.O., 1998. Applied Regression Analysis: A Research Tool, 2nd ed. Wadsworth & Brooks/Cole, Pacific Grove, CA

Russell, S.J., Norvig, P., 2002. Artificial Intelligence: A Modern Approach, 2nd edition. Prentice Hall, Englewood Cliffs, New Jersey 07632

Stearns, S.D., 1976. On selecting features for pattern classifiers. Proceedings of the 3rd International Joint Conference on Pattern Recognition, IEEE, pp. 71–75.

Todorovski, L. et al., 2004. Inducing polynomial equations for regression. Lecture notes in computer science, Lecture notes in artificial intelligence, 3201, Springer, Berlin, pp. 441-452.