CIplus
Band 4/2016

Building Ensembles of Surrogate Models by Optimal Convex Combination

Martina Friese, Thomas Bartz-Beielstein, and Michael Emmerich

Technology
Arts Sciences
TH Köln

BUILDING ENSEMBLES OF SURROGATE MODELS BY OPTIMAL CONVEX COMBINATION

Martina Friese and Thomas Bartz-Beielstein
SPOTSeven Lab, TH Köln
Steinmüllerallee 1, 51643 Gummersbach, Germany
{martina.friese|thomas.bartz-beielstein}@th-koeln.de

Michael Emmerich
LIACS, Leiden University
Niels Bohrweg 1, 2333CA Leiden, The Netherlands
m.t.m.emmerich@liacs.leidenuniv.nl

Abstract When using machine learning techniques for learning a function approximation from given data it is often a difficult task to select the right modeling technique. In many real-world settings is no preliminary knowledge about the objective function available. Then it might be beneficial if the algorithm could learn all models by itself and select the model that suits best to the problem. This approach is known as automated model selection. In this work we propose a generalization of this approach. It combines the predictions of several into one more accurate ensemble surrogate model. This approach is studied in a fundamental way, by first evaluating minimalistic ensembles of only two surrogate models in detail and then proceeding to ensembles with three and more surrogate models. The results show to what extent combinations of models can perform better than single surrogate models and provides insights into the scalability and robustness of the approach. The study focuses on multi-modal functions topologies, which are important in surrogate-assisted global optimization.

Keywords: Function Approximation, Surrogate Models, Model Selection, Ensemble Methods, Global Optimization

1. Introduction

Surrogate models are mathematical functions that, basing on a sample of known function values, approximate the behavior of the original function, while being cheaper in terms of evaluation. In the field of

optimization on expensive objective functions it is state of the art to use surrogate models to get an idea of the objective function landscape with lesser evaluations of the objective function. Expert systems like SPOT [1] come with a large variety of models that has to be chosen from when initiating an optimization process. The choice of the right model determines the quality of the the optimization process.
Often expert knowledge is needed to decide which model to select for a given problem. If there is no preliminary knowledge about the objective function it might be beneficial if the algorithm could learn all by itself which model suits best to the problem. This can be done by evaluating different models on test data a priori and using a statistical model selection approach to select the most promising model.
Some occurrences imply that there might also be a benefit in linearly combining predictors from several models into a more accurate predictor. In Fig. 1 such an occurrence is happening. Predictions with two different (Kriging) models are shown and results obtained by a convex combination of the predictors of these models. Different errors seem to be compensated by the combined model's predictions.
Such occurrences indicate that a predictor based on a single modeling approach is not always the best choice. On the other hand, complicated expressions based on multiple predictors might not be a good choice, either, due to overfitting and lack of transparency. Using convex combinations of predictors from available models seems to be a 'smart' compromise. Given surrogate models $\hat{y}_i : \mathbb{R}^d \to \mathbb{R}, i = 1, \ldots, s$, by a *convex combination of models* we understand a model given by $\sum_{i=1}^{s} \alpha_i \hat{y}$ with $\sum \alpha_i = 1$ and $\alpha_i \geq 0, i = 1, \ldots, s$. Finding an optimal convex combination of models can be viewed as a generalization of model selection. The selection of a single model is a special case with only one positive coefficient and the other coefficients zero.
This paper investigates the idea of using convex combinations of predictions of different models (*model mixtures*) to gain a more accurate predictions. Focusing on the predictions rather than implementation specialties when combing models gives us the ability to combine models without further considering the type of the model, making the approach very flexible. The main research questions are:

(Q-1) Can convex combinations of predictors improve as compared to (single) model selection?
(Q-2) Given the answer is positive, what are explanations of the observed behavior?
(Q-3) How can a system be build that finds the optimal convex combination of predictions on training data?

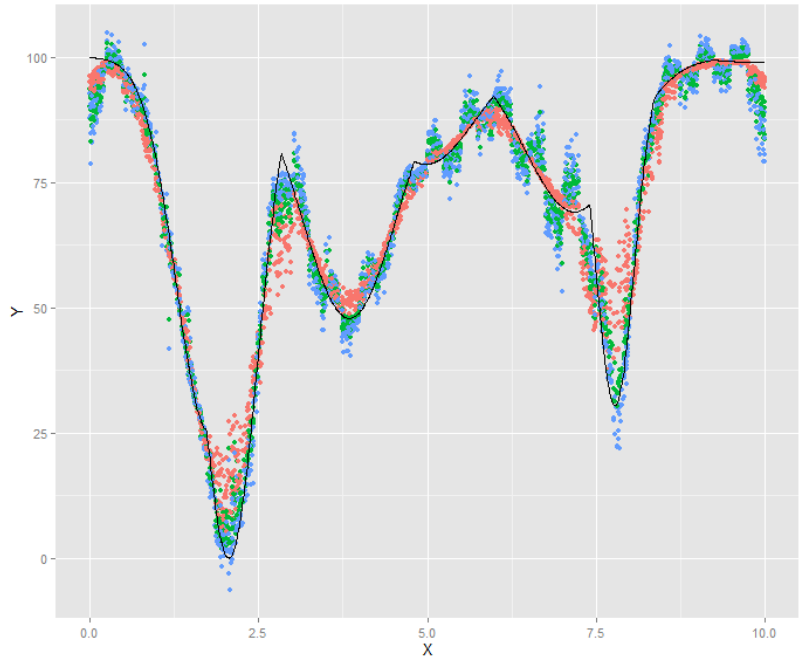

Figure 1: *The black line marks the actual objective function value. The dots show the results obtained in a leave-one-out cross-validation. Blue and red dots mark the predictions of single models. The green dots shows predictions obtained with an optimal linear combination of the two predictors.*

In order to answer these questions, detailed empirical studies are conducted, starting from simple examples and advancing to more complex ones. To improve readability, this paper follows a non-standard structure, where the discussion of experimental results follows directly the introduction of the modeling extensions.
The paper is structured as follows: Section 2 discusses the general approach and related work. Section 3 provides technical preliminaries for the subsequent experiments. Section 4 introduces the idea of model mixtures and explores binary model mixtures. Section 5 provides a more detailed analysis of binary model mixtures. Section 6 extends the analysis to ternary model mixtures, and Section 7 provides first results and techniques for enabling mixtures of a larger number of models. Section 8 discusses the main results and future research directions.

2. General Approach and Related Work

To base a decision or build a prediction from multiple opinions is common practice in our everyday live. It happens in a democratic government, or when in TV shows the audience is asked for help. One also might use it when we try to build an opinion on a topic that is new to us. Naturally, such tools already found their way into statistical prediction and machine learning.

In statistics and machine learning an *ensemble* is a prediction model from several prediction models. A comprehensive introduction to ensemble-based approaches in decision making is given in [6] and [4]. Generally, there are two groups of ensemble approaches: the first group's approaches, the so-called *single-evaluation* approaches, only choose and build one single model, whereas the second group's approaches, the so-called *multi-evaluation* approaches, build all models, and use the derived information to decide which output to use. For each of these two approaches, several model selection strategies can be implemented. Well-known strategies are:

- *Round robin* and *randomized choosing* are the most simplistic implementations of ensemble-based strategies. In the former approach, the models are chosen in a circular order independent of their previously achieved gain. In the latter approach, the model to be used in each step is selected randomly from the list of available models. The previous success of the model is not a decision factor.
- *Greedy strategies* choose the model that provided the best function value so far, while the SoftMax strategy uses a probability vector, where each element represents the probability for a corresponding model to be chosen [8]. The probability vector is updated depending on the reward received for the chosen models.
- *Ranking strategies* try to combine the responses of all meta models to one response, where all meta models contributed to, rather than to choose one response.
- *Bagging* combines results from randomly generated training sets and can also be used in function approximation, whereas
- *Boosting* combines several weak learners to a strong one in a stochastic setting.
- *Weighted averaging* approaches do not choose a specific model's result but rather combine it by averaging. Since bad models should not deteriorate the overall result, a weighting scheme is introduced. Every model's result for a single design point is weighted by its overall error, the sum over all models yields the final value assigned to the design point. In *stacking*, several trained models are combined and trained

again by a stacking algorithm. A typical example of a successful weighted average model are Random Forests [3].

Convex combinations of surrogate models used in this paper can be viewed as a special case of weighted averaging models, albeit we propose here global optimization instead of re-training for finding the *best convex combination* of models. Moreover, the analysis in this paper aims for transparent presentation of results using mixture analysis and focuses on multimodal function approximation, which is an important application in surrogate-assisted global optimization.

3. Preliminaries

3.1 Surrogate Models

By a surrogate model, we will understand here a function $\hat{y} : \mathbb{R}^d \to \mathbb{R}$ that is an approximation to the original function $y : \mathbb{R}^d \to \mathbb{R}$, and learned from a finite set of evaluations of the original function. A typical application of surrogate models is to provide a fast approximations of functions that are expensive to evaluate, for instance functions based on costly computer simulations. Kriging surrogate models were used in our study. A set of three different kernels was used to implement the ensemble strategies. Following the definitions from [7], the correlation models can be described as follows. We consider stationary correlations of the form

$$\mathcal{R}(\theta, w, x) = \prod_{j=1}^{n} \mathcal{R}(\theta_j, w_j - x_j).$$

The first model uses the *exponential* kernel

$$\mathcal{R}(\theta, w, x) = \exp(-\theta_j |w_j - x_j|),$$

the second model uses an *gaussian* kernel

$$\mathcal{R}(\theta, w, x) = \exp(-\theta_j |w_j - x_j|^2),$$

whereas the third model is based on the *spline correlation* function $\mathcal{R}(\theta, w, x) = \zeta(\theta_j |w_j - x_j|)$ with

$$\zeta(\epsilon_j) = \begin{cases} 1 - 15\epsilon_j^2 + 30\epsilon_j^3 & \text{for} \quad 0 \le \epsilon \le 0.2 \\ 1.25(1 - \epsilon_j)^3 & \text{for} \quad 0.2 < \epsilon_j < 1 \\ 0 & \text{for} \quad \epsilon_j \ge 1. \end{cases}$$

The variables ϵ and θ are hyperparameters estimated by likelihood maximization.

Table 1: *Gaussian landscape generator options*

Parameter	*Description*	*Value*
n	Dimension	2 - 10
m	Number of peaks	10 - 40
l	Lower bounds of the region, where peaks are generated	{0; 0}
u	Upper bounds of the region, where peaks are generated	{5; 5}
max	Max function value	100
t	Ratio between global and local optima	0.8

3.2 Objective Functions

For generating *test functions* we used the *Max-Set of Gaussian Landscape Generator* (MSG) [5], which can be used to set up problem instances for continuous, bound-constrained optimization problems. It uses the maximum of m weighted Gaussian functions

$$G(x) = \max_{i \in 1,2,\ldots,m} (w_i g_i(x)),$$

where $g : \mathbb{R}^n \to \mathbb{R}$ denotes an n-dimensional Gaussian function

$$g(x) = \left(\frac{\exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right)}{(2\pi)^{n/2}|\Sigma|^{1/2}} \right)^{1/n},$$

μ is an n-dimensional vector of means, and Σ is an $(n \times n)$ covariance matrix. The mean of each Gaussian corresponds to an optimum on the landscape and the location of all optima is known. The global optimum is the one with the largest value. For the generation of the objective function the `spotGlgCreate` method of the SPOT package has been used. Implementation details are presented in [2]. The options used for our experiments are shown in Table 1. With the parameter n the dimension of the objective function is specified. The lower and upper bounds (l and u, respectively) specify the region where the peaks are generated. The value max specifies the function value of the global optimum, while the maximum function value of all other peaks is limited by t, the ratio between the global and the local optima.

4. Binary Ensembles

This Section analyses models which combine only two models. Convex combinations of models will be referred to as ensemble models, while the

original models will be referred to as base models. We focus on positive weights, since we do not want to select models that make predictions which are anti-correlated with the results.
A sample of points (design) is evaluated on the objective function (MSG, for parameters see Table 1). For the sampling of the points a latin hypercube design featuring 40 design points is generated. The two base models are Kriging with exponential correlation function (referred to as a) and gaussian correlation function (referred to as b). Both base models are fitted to the data and then asked to do a prediction on the testdata. The predictions $\hat{y}$ of the ensemble models are calculated as linear combinations of the predictions of the base models.
Given a weight α_i, where $\alpha_i \in \{0.0, 0.1, 0.2, ..., 0.9, 1.0\}$, the ensemble models can be defined as the linear combinations of the models a and b as follows:

$$\hat{y}_n = \alpha_n \times \hat{y}_a + (1 - \alpha_n) \times \hat{y}_b \tag{1}$$

The models are evaluated by calculating the root mean squared error (RMSE) of the predictions made during a leave-one-out cross-validation on the 40 design points.
Since randomness has been induced into the experiment by using the Latin hypercube design, the evaluation process has been repeated 50 times. With each model returning one prediction for each design point in every repetition this results in a total of $2,000$ prediction values (40 design points $\times$ 50 repetitions) for each model.

To get a first quick insight into the result data, for each repetition the rankings of the RMSE's have been calculated. The models with $\alpha = 0.6$, $\alpha = 0.8$ and $\alpha = 0.9$ dominate this comparison, each performing best 8 out of 50 times. The base models, a and b, performed best only in four respectively two cases out of 50. Never an ensemble model with positive weights was performing worst.
In order to achieve some comparability between the RMSE's of different repetitions all RMSE's have been repetition-wise scaled to values between zero and one, so that the scaled RMSE of the best model in one repetition is always zero and the scaled RMSE of the worst model for one repetition is always 1.0. Figure 2 shows the boxplot over these scaled RMSE's. It can be seen that the model a (exponential) in most of the cases performs worst since its median value is one, only some outliers come closer to zero.
Model b (Gaussian) shows a larger variation in its performance. It has been the best- as well as the worst performing model each at least once. Its median and mean performances are average in comparison with all models evaluated.

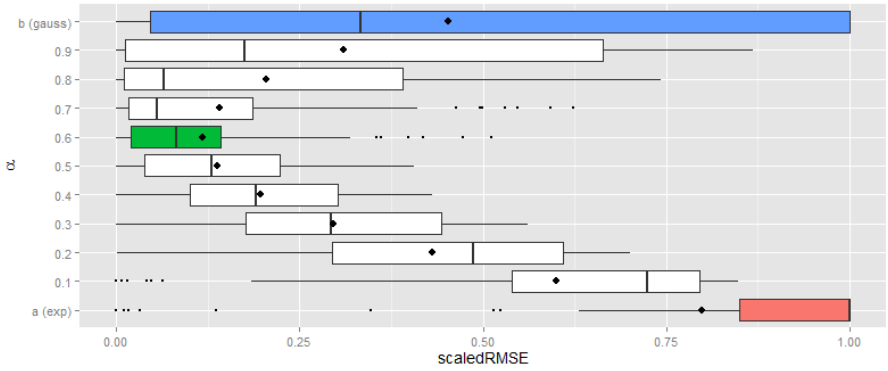

Figure 2: *Boxplot over the scaled RMSE's of all models. The models are defined by an α-weighted linear combination of the two base models. The results of the base models depicted on the outer rows and colored red (exponential kernel), respectively blue (Gaussian kernel). All linear combinations are drawn in between. The model combination chosen as best with $\alpha = 0.6$ is colored green. The mean value of each result bar is marked by a dot.*

A parabolic tendency can be seen in the performances. This indicates that a linear combinations of the models are indeed beneficial. Due to the convex combination of the predictor, a prediction by the ensemble model cannot be worse but it might be better than both base models. An ensemble can only be better, if one model overestimates and the other model underestimates the original function value. In the experiment this happens in 649 out of 2000 cases.

As a *consistent method for evaluating the performance and automatically choosing the best model* the following approach is proposed: Model-wise mean-, median- and 3rd quartile-values are calculated. The resulting values are ranked and the rankings summed up to one final ranking. The model that achieved the lowest value is recommended as best choice. In Figure 2 the model recommended as best choice by this method is colored green.

5. Detailed Analysis on Transparent Test Cases

It can clearly be stated that for this first experiment setup the combination of two models is beneficial for the overall prediction. In this section we're going to have a closer look at possible explanations for the successful result. Are there problem features that encourage using ensembles and is this result generalizable?

5.1 2D Experiment Setup and Analysis

The nature of the combination method we used suggests, that the use of an ensemble built by linear combination is beneficial in cases when one of the base models underestimates the objective function value while the other overestimates it.
Figure 3 depicts additional results on the experiment setup carried out in Sec. 4. The boxplot in the middle on the left column is the already known one. The contour plot in the upper left shows the objective

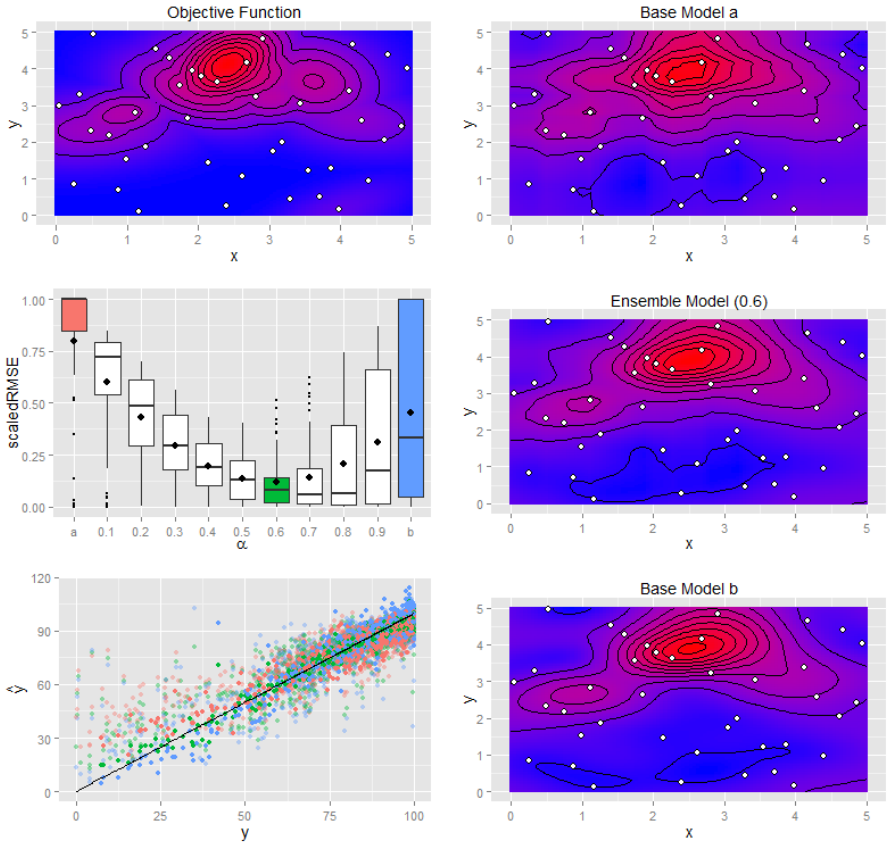

Figure 3: *Additional results on the experiment carried out in Sec. 4. The contour plots show the actual objective function and the fits of the colored models at the best choice's mean performance. In the lower left the predictions $\hat{y}$ are plotted against the actual function values y.*

function while the contour-plots in the right column show the fits of the base models and the best choice model. Since during the experiment each model has been fitted 50 times, the fits shown here relate to the

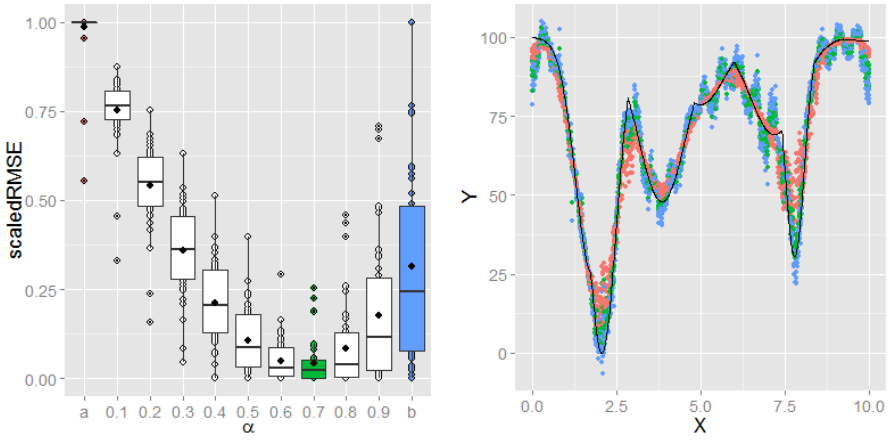

Figure 4: *Results on a 1D objective function. The boxplot shows the scaled RMSE's of the models over the experiments 50 repetitions. The α value defines the weight for the linear combination. The ensemble obtained by a linear combination with $\alpha = 0.7$, here colored green, is suggested best for this experiment setup. On the right hand side all predictions done during the leave-one-out cross validation for the base models and the best model are plotted against the objective function.*

repetition where the best choice model showed it's mean performance. In these plots the white dots mark the points of the design used to evaluate the models fits. Variations in these plots are visible, but the benefit of the ensemble model remains invisible.
The plot in the lower left shows results in more detail. Here the predictions $\hat{y}$ are plotted against the actual function values y. The black line marks the objective function values y. Only predictions $\hat{y}$ for the same parameter set (x_1, x_2), where the predictions $\hat{y}$ of the base models span the function value y, are plotted with full opacity. Prediction sets, where both base models either over- or underestimated the actual function value are drawn only with a light opacity. In this experiment 649 out of 2000 prediction sets span the actual objective function value.

5.2 1D Experiment Setup and Analysis

Since the 2D experiment setup from Sec. 4 does not allow for an easy analysis of the results, the experiment has been redone on a 1D objective function to allow for a better understanding of the underlying process. The only change in the experimental setup that has been made is the dimension of the underlying objective function, which has been set to 1. The main results of this second experiment setup are depicted in Figure 4. The boxplot on the left hand side shows the scaled RMSE's for all

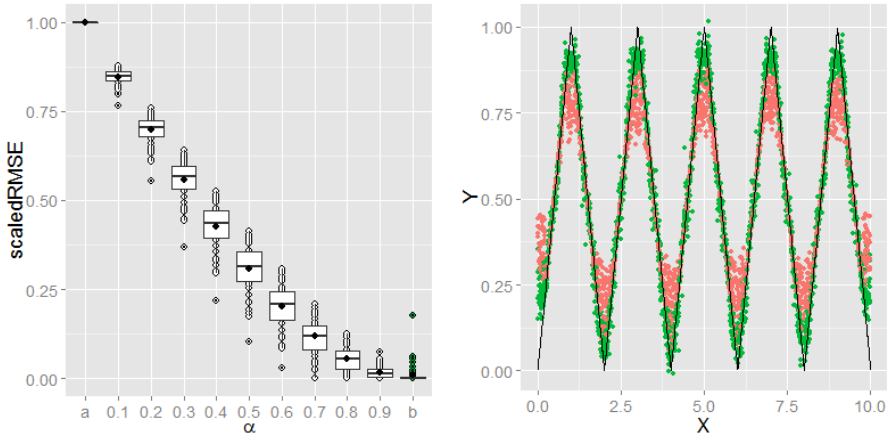

Figure 5: *Results on a triangle objective function. Left hand side plot shows the scaled RMSE's. The α value defines the weight for the linear combination. Here the base model b ($\alpha = 1.0$) is chosen best. On the right hand side all predictions done during the leave-one-out cross validation for the base models and the best model are plotted against the objective function. Since model b has been chosen best it is colored green.*

models. Applying the rule defined in Sec. 4 names the model obtained by a linear combination with $\alpha = 0.7$ as the best choice. The plot on the right hand side shows only the performance of the best choice model and the base models. Each dot marks a single prediction. As can be seen in the plot, the predictions of the model a (exponential), marked by red dots,seem to smooth the objective function - straight segments are well met while curved segments are smoothed out.
The predictions of the model b (gaussian), marked by the blue dots show signs of overfitting. Again straight segments are well met but when approaching local extrema the predictions start to oscillate. So the linear combination of both predictions averages positive as well as negative outliers of base models. This seems to provide some benefit to the overall experiment outcome.

Since the curves and corners in the objective function seem to make the game here, two additional experiments are set up. For these experiments two functions are specified featuring corners that are not continuous differentiable. For one experiment a triangle function is used for another a piecewise assembled function. Whereas looking at the results on the piecewise assembled function we again find a strong parabolic tendency in the boxplot. Both base models have a rather large variance in their

performance. The ensemble model marked as best choice has a smaller variance and performed better than the base models in nearly all cases.

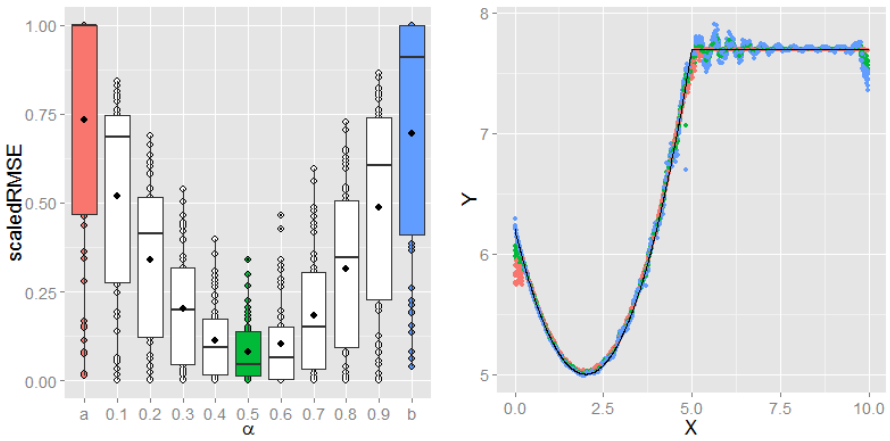

Figure 6: *Results on a piecewise assembled objective function. Left hand side plot shows the scaled RMSE's. The α value defines the weight for the linear combination. The ensemble obtained by a linear combination with $\alpha = 0.5$, here colored green, is suggested best for this experiment setup. On the right hand side all predictions done during the leave-one-out cross validation for the base models and the best model are plotted against the objective function.*

Figures 5 and 6 show the results of these experiments. Looking at the experiment results featuring the triangle objective function, the boxplot shows a clear tendency towards base model b and in both plots only the results of two models are colored. Here base models b clearly outperformed all other models und thus was chosen best.

6. Ternary Ensembles

Next, the experiments are extended to a larger scale: The dimensionality of the objective function is increased and three base models are combined. As before Kriging models with different kernels are used, but now a third model using the spline correlation function is added.

$$\alpha_n, \beta_n, \gamma_n \in \{0.0, 0.1, 0.2, ..., 0.9, 1.0\}, \quad \alpha_n + \beta_n + \gamma_n = 1 \qquad (2)$$

For the linear combination of three base models three weights are needed, that sum up to one as specified in (2). With a step size of 0.1 for the linear combinations this results in 66 models.
Figure 7 shows the results of the first experiment using three base models. The only change that has been made to the original experiment

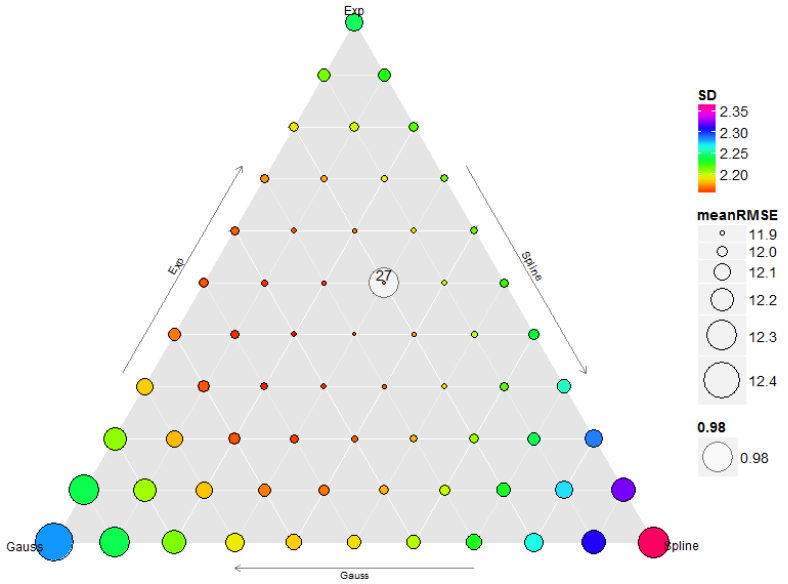

Figure 7: *The plot shows the results of the experiment set up with three base models. Each circle depicts the performance results for one model . The three base models are located on the corners of the triangle, models gained by linear combinations of only two models are located on the outer border. Circles on the inner area of the area show the results for models that were gained by linear combinations of all three base models. The size of the circles denotes the mean RMSE value, the color the standard deviation. The model proposed as best choice is marked by an additional white circle.*

setup, besides the number of base models, is the dimension n of the objective function and the number of peaks m generated in the Gaussian landscape. As a first step towards objective functions of higher complexity, the dimension of the objective function has been set to 4. But this change alone is not sufficient to gain a larger complexity, since without adjusting the number of Gaussian components used for generating the objective function, it rather gets less complex. Thus the number of Gaussians used has been adjusted to ten times the Dimension. With the points getting smaller when approaching the center of the triangle, it can be stated, that again it is beneficial to do a linear combination of the base models.

7. Scaling-up to Multiple Models

Up to this point only experiments with up to three models have been carried out, but the underlying goal is to evolve a system that is able to handle quite a large set of available base models. But at this point quickly another approach is needed, since the number of possible linear combinations between a higher number of base models grows exponentially.

$$f(s,n) = 1 + \sum_{s^*=1}^{s-1} f(s - s^*, n-1), \quad f(s,1) = 1, f(1,n) = n \qquad (3)$$

The relation between number of models, the step width for the linear combinations and the resulting number of linear combinations can be expressed as function of s the reciprocal of the step width and n the number of models as defined in (3). Using three base models and a step width of 0.1 as defined in (2) this results in $f(10,3) = 66$ linear combinations that have to be taken account of. Now thinking of combinations of 10 base models already results in $f(10,10) = 92378$ linear combinations. The complexity of the search space, when increasing the number of models, quickly gets too large to do a complete evaluation of all possible linear combination with a fixed step width of 0.1. Keeping in mind that, looking at previous results, the function that describes the performance of the models built by linear combinations up to this point only showed unimodal characteristics, which seems to be expectable due to its nature. We expect the function to show this characteristic also when combining larger number of models.

Thus at this point, instead of a complete evaluation of all linear combinations, an optimization step has been implemented to find the best combination. The allowed weights have been restricted to a precision of two decimal places. Since the area around the optimum tends to build a plateau, this reduces the possible search space without pruning the possible best solution. In behalf of comparability, the experiment setup here is exactly the same as the one used in Sec. 6. Only the process itself changed. Beforehand all possible linear combinations have been evaluated. Now, only the base models have been evaluated, all other models were only evaluated during the evolutionary process. We also stuck to the method used by the (1+1)-ES of comparing the offspring only to the parent rather than to the whole population as we did it before.

For the mutation of the weights vector $\vec{v} = (\alpha, \beta, \gamma)^T$ three random samples of a normal distribution function with standard deviation of 0.16 have been drawn and added to the weights vector. Since this alone

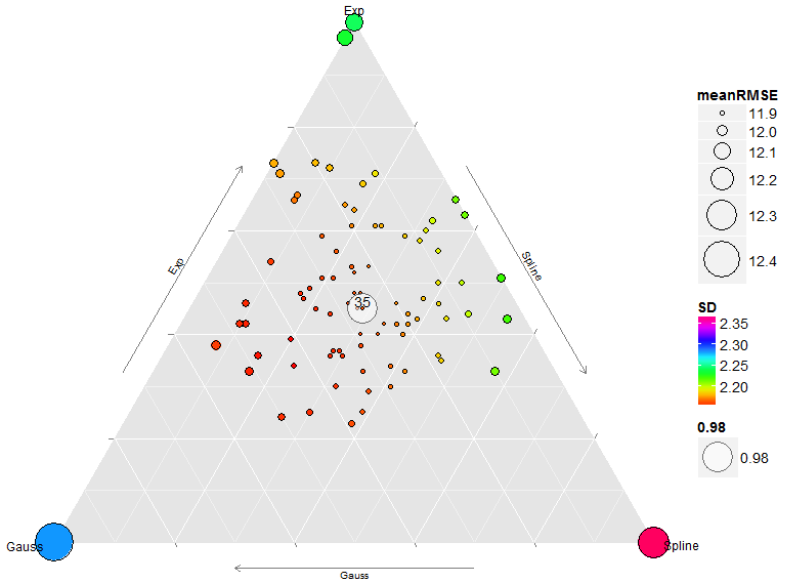

Figure 8: *The plot shows the results of the same experiment setup as presented in Sec. 6. The optimal linear combination has been searched with a simple (1+1)-Evolution Strategy with 1/5th success rule. Again, each circle depicts the performance results for one model. The three base models are located on the corners of the triangle, models gained by linear combinations of only two models are located on the outer border. Circles on the inner area of the area show the results for models that were gained by linear combinations of all three base models. The size of the circles denotes the mean RMSE value, the color the standard deviation. The model proposed as best choice is marked by an additional white circle.*

does not meet the requirements needed for a valid weights vector, the resulting vector has been adjusted in 3 steps:

1) If $\min(\alpha, \beta, \gamma) < 0$ then $\vec{v} := \vec{v} - \min(\alpha, \beta, \gamma)$,
2) $\vec{v} := \vec{v}/(\alpha + \beta + \gamma)$,
3) Round the values α, β, γ to two decimal places so, that $\alpha + \beta + \gamma = 1$.

For this experiment we allowed a maximum of 100 individuals to be evaluated. Within these bounds already the 35th individual evaluated has been the best individual found in this run. Figure 8 depicts the results of this optimization step. As before, the best individual is marked by a white circle.

8. Discussion and Outlook

Reconsidering the research questions from Sec. 1, we can state that linear combinations of predictors can generate better results than model selection (Q-1). A system, which finds optimal linear combinations, was presented in Sec. 4. The corresponding experiments were extended to a larger scale in Sec. 6. The results from these experiments further support our statement, that combination of models leads to better results. Finally, in Sec. 7, we proposed a method to include even more base models to the system. For the same experiment setup as used before, a solution of comparable performance quality has been found, with even lesser number of ensemble model evaluations. With this method the foundation has been created for a larger system including all available models.

Although research question (Q-3) could be answered positively, a complete answer to question (Q-2) could not be given in this study. Explanations of the observed behavior require further research. Ideas and questions that were not investigated so far include:

- Experiments featuring more base models, also including other types of models.
- Extensive analysis of the influence of objective function attributes on the experiment outcome. The results of Sec. 5.2 suggest, that particularly piecewise assembled functions might be of special interest.
- Studies also allowing other operations than simple linear combinations only.
- Conception of a procedure that includes our method of ensemble building into an sequential optimization process.

Summarizing, this preliminary study presents valuable and new findings in the field of ensemble-based modeling. We developed a smart and simple strategy strategy for combining different modeling approaches. It uses a (linear) combination of the predicted values and is easily applicable in many modeling situations where several models are available. Especially, if the user does not know, which model to choose, a linear combination might be a promising approach. The weights in the linear model can shed some light on the relevance of certain models and illustrate, which model is active. Ternary plots (as shown in Figure 7) can be used to illustrate the progress of the optimization process. However, since determination of optimal weights in the linear model is a non-linear optimization problem, we cannot guarantee the optimality of the proposed weights. All in all, this experimental study presents some important findings about the behavior of an ensemble-based approach that defines an interesting direction of research.

References

[1] T. Bartz-Beielstein. Spot: An R package for automatic and interactive tuning of optimization algorithms by sequential parameter optimization. Technical Report 05/10, Research Center CIOP (Computational Intelligence, Optimization andData Mining), Cologne University of Applied Science, Faculty of Computer Science and Engineering Science, 2010. Comments: Related software can be downloaded from http://cran.r-project.org/web/packages/SPOT/index.html.

[2] T. Bartz-Beielstein. How to Create Generalizable Results. In J. Kacprzyk and W. Pedrycz, editors, Springer Handbook of Computational Intelligence, pages 1127–1142. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[3] L. Breiman. Random forests. Mach. Learn., 45(1):5–32, Oct. 2001.

[4] M. Friese, M. Zaefferer, T. Bartz-Beielstein, O. Flasch, P. Koch, W. Konen, and B. Naujoks. Ensemble-Based Optimization and Tuning Algorithms. In F. Hoffmann and E. Hüllermeier, editors, Proceedings 21. Workshop Computational Intelligence, pages 119–134. Universitätsverlag Karlsruhe, 2011.

[5] M. Gallagher and B. Yuan. A general-purpose tunable landscape generator. IEEE Trans. Evolutionary Computation, 10(5):590–603, 2006.

[6] R. Polikar. Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 6(3):21–45, 2006.

[7] J. S. Søren N. Lophaven, Hans Bruun Nielsen. Dace - a matlab kriging toolbox. Technical report, Technical University of Denmark, 2002.

[8] R. S. Sutton and A. G. Barto. Introduction to Reinforcement Learning. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

CIplus
Band 4/2016

Building Ensembles of Surrogate Models by Optimal Convex Combination

Martina Friese
Thomas Bartz-Beielstein
Technische Hochschule Köln

Michael Emmerich
LIACS, Leiden University

März 2016

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Technology
Arts Sciences
TH Köln