Title: Surrogate model

URL: https://en.wikipedia.org/wiki/Surrogate_model

PageID: 6500531

Categories: Category:Design of experiments, Category:Machine learning, Category:Mathematical modeling, Category:Numerical analysis, Category:Scientific models

-----

A surrogate model is an engineering method used when an outcome of interest cannot be easily measured or computed, so an approximate mathematical model of the outcome is used instead. Most engineering design problems require experiments and/or simulations to evaluate design objective and constraint functions as a function of design variables. For example, in order to find the optimal airfoil shape for an aircraft wing, an engineer simulates the airflow around the wing for different shape variables (e.g., length, curvature, material, etc.). For many real-world problems, however, a single simulation can take many minutes, hours, or even days to complete. As a result, routine tasks such as design optimization , design space exploration , sensitivity analysis and "what-if" analysis become impossible since they require thousands or even millions of simulation evaluations.

One way of alleviating this burden is by constructing approximation models, known as surrogate models , metamodels or emulators , that mimic the behavior of the simulation model as closely as possible while being computationally cheaper to evaluate. Surrogate models are constructed using a data-driven, bottom-up approach. The exact, inner working of the simulation code is not assumed to be known (or even understood), relying solely on the input-output behavior. A model is constructed based on modeling the response of the simulator to a limited number of intelligently chosen data points. This approach is also known as behavioral modeling or black-box modeling, though the terminology is not always consistent. When only a single design variable is involved, the process is known as curve fitting .

Though using surrogate models in lieu of experiments and simulations in engineering design is more common, surrogate modeling may be used in many other areas of science where there are expensive experiments and/or function evaluations.

Goals

The scientific challenge of surrogate modeling is the generation of a surrogate that is as accurate as possible, using as few simulation evaluations as possible. The process comprises three major steps which may be interleaved iteratively:

Sample selection (also known as sequential design, optimal experimental design (OED) or active learning )

Construction of the surrogate model and optimizing the model parameters (i.e., bias-variance tradeoff )

Appraisal of the accuracy of the surrogate.

The accuracy of the surrogate depends on the number and location of samples (expensive experiments or simulations) in the design space. Various design of experiments (DOE) techniques cater to different sources of errors, in particular, errors due to noise in the data or errors due to an improper surrogate model.

Types of surrogate models

Popular surrogate modeling approaches are: polynomial response surfaces ; kriging ; more generalized Bayesian approaches; [ 1 ] gradient-enhanced kriging (GEK); radial basis function ; support vector machines ; space mapping ; [ 2 ] artificial neural networks and Bayesian networks . [ 3 ] Other methods recently explored include Fourier surrogate modeling [ 4 ] [ 5 ] and random forests . [ 6 ]

For some problems, the nature of the true function is not known a priori , and therefore it is not clear which surrogate model will be the most accurate one. In addition, there is no consensus on how to obtain the most reliable estimates of the accuracy of a given surrogate. Many other problems have known physics properties. In these cases, physics-based surrogates such as space-mapping based models are commonly used. [ 2 ] [ 7 ]

Invariance properties

Recently proposed comparison-based surrogate models (e.g., ranking support vector machines ) for evolutionary algorithms , such as CMA-ES , allow preservation of some invariance properties of surrogate-assisted optimizers: [ 8 ]

Invariance with respect to monotonic transformations of the function (scaling)

Invariance with respect to orthogonal transformations of the search space (rotation)

Applications

An important distinction can be made between two different applications of surrogate models: design optimization and design space approximation (also known as emulation).

In surrogate model-based optimization, an initial surrogate is constructed using some of the available budgets of expensive experiments and/or simulations. The remaining experiments/simulations are run for designs which the surrogate model predicts may have promising performance. The process usually takes the form of the following search/update procedure.

Initial sample selection (the experiments and/or simulations to be run)

Construct surrogate model

Search surrogate model (the model can be searched extensively, e.g., using a genetic algorithm , as it is cheap to evaluate)

Run and update experiment/simulation at new location(s) found by search and add to sample

Iterate steps 2 to 4 until out of time or design is "good enough"

Depending on the type of surrogate used and the complexity of the problem, the process may converge on a local or global optimum , or perhaps none at all. [ 9 ]

In design space approximation, one is not interested in finding the optimal parameter vector, but rather in the global behavior of the system. Here the surrogate is tuned to mimic the underlying model as closely as needed over the complete design space. Such surrogates are a useful, cheap way to gain insight into the global behavior of the system. Optimization can still occur as a post-processing step, although with no update procedure (see above), the optimum found cannot be validated.

Surrogate modeling software

Surrogate Modeling Toolbox (SMT: https://github.com/SMTorg/smt ) is a Python package that contains a collection of surrogate modeling methods, sampling techniques, and benchmarking functions. This package provides a library of surrogate models that is simple to use and facilitates the implementation of additional methods. SMT is different from existing surrogate modeling libraries because of its emphasis on derivatives , including training derivatives used for gradient -enhanced modeling, prediction derivatives, and derivatives with respect to the training data. It also includes new surrogate models that are not available elsewhere: kriging by partial-least squares reduction and energy-minimizing spline interpolation . [ 10 ]

Python library SAMBO Optimization supports sequential optimization with arbitrary models, with tree-based models and Gaussian process models built in. [ 11 ]

Surrogates.jl is a Julia packages which offers tools like random forests, radial basis methods and kriging.

Surrogate-Assisted Evolutionary Algorithms (SAEAs)

SAEAs are an advanced class of optimization techniques that integrate evolutionary algorithms (EAs) with surrogate models. In traditional EAs, evaluating the fitness of candidate solutions often requires computationally expensive simulations or experiments. SAEAs address this challenge by building a surrogate model, which is a computationally inexpensive approximation of the objective function or constraint functions.

The surrogate model serves as a substitute for the actual evaluation process during the evolutionary search. It allows the algorithm to quickly estimate the fitness of new candidate solutions, thereby reducing the number of expensive evaluations needed. This significantly speeds up the optimization process, especially in cases where the objective function evaluations are time-consuming or resource-intensive.

SAEAs typically involve three main steps: (1) building the surrogate model using a set of initial sampled data points, (2) performing the evolutionary search using the surrogate model to guide the selection, crossover, and mutation operations, and (3) periodically updating the surrogate model with new data points generated during the evolutionary process to improve its accuracy.

By balancing exploration (searching new areas in the solution space) and exploitation (refining known promising areas), SAEAs can efficiently find high-quality solutions to complex optimization problems. They have been successfully applied in various fields, including engineering design, machine learning, and computational finance, where traditional optimization methods may struggle due to the high computational cost of fitness evaluations. [ 12 ] [ 13 ]

See also

Linear approximation

Response surface methodology

Kriging

Radial basis functions

Gradient-enhanced kriging (GEK)

OptiY

Space mapping

Surrogate endpoint

Surrogate data

Fitness approximation

Computer experiment

Conceptual model

Bayesian regression

Bayesian model selection

References

Further reading

Queipo, N.V., Haftka, R.T., Shyy, W. , Goel, T., Vaidyanathan, R., Tucker, P.K. (2005), " Surrogate-based analysis and optimization ," Progress in Aerospace Sciences, 41, 1–28.

D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, K. Crombecq, (2010), " A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design ," Journal of Machine Learning Research, Vol. 11, pp. 2051–2055, July 2010.

T-Q. Pham, A. Kamusella, H. Neubert, " Auto-Extraction of Modelica Code from Finite Element Analysis or Measurement Data ," 8th International Modelica Conference, 20–22 March 2011 in Dresden.

Forrester, Alexander, Andras Sobester, and Andy Keane, Engineering design via surrogate modelling: a practical guide , John Wiley & Sons, 2008.

Bouhlel, M. A. and Bartoli, N. and Otsmane, A. and Morlier, J. (2016) " Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction ", Structural and Multidisciplinary Optimization 53 (5), 935–952

Bouhlel, M. A. and Bartoli, N. and Otsmane, A. and Morlier, J. (2016) " An improved approach for estimating the hyperparameters of the kriging model for high-dimensional problems through the partial least squares method ", Mathematical Problems in Engineering

External links

Matlab code for surrogate modelling

Matlab SU rrogate MO deling Toolbox – Matlab SUMO Toolbox

Surrogate Modeling Toolbox -- Python