Metamodels for mixed variables by multiple kernel regression

Manuel Herrera, Rajan Filomeno Coelho[†]

[†]BATir Department, Université libre de Bruxelles
Avenue Franklin Roosevelt, 50 (CP 194/2), 1050 Brussels (Belgium)
{mherrera,rfilomen}@ulb.ac.be

Keywords: surrogate models, support vector regression, kernel methods, mixed variables, truss structures

ABSTRACT

Metamodel-assisted optimization has greatly improved the design of mechanical components and civil engineering structures, due to their capacity to address physically complex problems through the use of inexpensive interpolation or regression models [1]. However, a majority of existing surrogate models encountered in the literature focus on continuous inputs, viz. they do not take explicitly into account discrete, integer, or categorical values, although versatile practical engineering problems also involve non-continuous parameters. In particular, categorical variables can represent any non-numerical data, like a performance assessment ('low', 'medium', 'high'), or the choice of a material ('steel', 'titanium', 'aluminum'); in the former case, they are said to be ordered (*ordinal* variables), while they are unordered (*nominal* variables) in the latter case [2, 3].

Preliminary studies by the authors [4]–based on the development of moving least squares adapted to mixed (continuous and nominal) variables–have demonstrated their efficiency for a low number of nominal variables and a limited number of *attributes* (i.e. the possible values for the nominal variables). However, these approaches do not infer any *a priori* relationship between the inputs (e.g. in a structural design problem, the geometry of a beam cross-section: 'square', 'circle', 'I',...) and the outputs (e.g. the maximum deflection of the beam at mid-span): all attributes are implicitly considered as equally distant in the design space, while in practice clusters of attributes could be determined according to their corresponding influence on the outputs.

Therefore, the aim of this work is to propose a *multiple kernel learning* (MKL) alternative to develop efficient surrogate models which can handle continuous and categorical variables by a number of mapping functions combined. Common kernel-based learning methods use an implicit mapping of the input data into a high dimensional feature space defined by a kernel function, i.e., a function returning the inner product $\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle$ between the images of two data points $\boldsymbol{x}, \boldsymbol{x}'$ in the feature space. The learning then takes place in the feature space and the algorithm can be expressed so that the data points only appear inside dot products with other points. This is often referred to as the "kernel trick" [5].

Kernel selection is very important because its choice is highly dependent on the nature of the input data. Furthermore, we might have access to multiple distinct types of input data when performing regression (or classification) tasks. In such situations it may be difficult to determine the single most suitable kernel and the performance of the method could suffer as a result. Instead we could try to find the optimal combination of a set of candidate kernels, where each of the kernels represents a different type of data. The proposal here is to apply appropriate kernel functions to generate individual kernel matrices for the different types of data. These kernels can then be combined with a weighted summation and used as training data for a classical support vector regression (SVR). This approach is referred to as multiple kernel learning (MKL) [6] and the problem of combining kernels is the problem of data integration, which can be addressed by a gradient descent algorithm.

The MKL regression proposed here has been successfully validated on a set of six analytical mixed-variate benchmark functions, and applied to the design of a six-storey rigid frame [7] characterized by continuous (cross-section diameters and thicknesses) and nominal (geometry of the cross-sections) variables. The nominal variables have been represented by a so-called dummy coding. In all cases, MKL outperforms the other approaches proposed in the literature to deal with categorical variables (ordinary and moving least squares). The proposed methodology seems promising to deal efficiently with different kinds of information or data types.

The full methodology summarized here–along with the corresponding numerical results and discussion–will be presented thoroughly during the oral presentation.

References

[1] P Breitkopf and R Filomeno Coelho, editors. *Multidisciplinary Design Optimization in Computational Mechanics*. ISTE/John Wiley & Sons, Chippenham, UK, April 2010. 1 volume, 549 pages.

[2] A Agresti. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York, 1996.

[3] B McCane and M H Albert. Distance functions for categorical and mixed variables. *Pattern Recognition Letters*, 29(7):986–993, 2008.

[4] R Filomeno Coelho. Extending moving least squares to mixed variables for metamodel-assisted optimization. In *6th European Congress on Computational Methods in Applied Sciences and Engineering – ECCOMAS 2012, Vienna, Austria, September 10-14*, 2012.

[5] B Schölkopf and A J Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, USA, 2001.

[6] G Lanckriet, N Cristianini, L El Ghaoui, P Bartlett, and M Jordan. Learning the kernel matrix with semi-definite programming. *Journal of Machine Learning Research*, 5:27–72, 2004.

[7] M Papadrakakis and N D Lagaros. Soft computing methodologies for structural optimization. *Applied Soft Computing*, 3:283–300, 2003.