MULTIOBJECTIVE MODEL SELECTION FOR NON-LINEAR REGRESSION TECHNIQUES

Luca Pasolli[1,2], Claudia Notarnicola[2], Lorenzo Bruzzone[1]

[1]Dep. of Information Engineering and Computer Science, University of Trento,
Via Sommarive, 14, I-38123 Trento Italy. e-mail: lorenzo.bruzzone@ing.unitn.it
[2]Institute for Applied Remote Sensing, EURAC Research
Viale Druso, 1 I-39100 Bolzano Italy.

1. INTRODUCTION AND MOTIVATION

The estimation of biophysical parameters from remotely sensed data represents an important and challenging research field within the remote sensing community. A possible and effective way to deal with this issue consists in adopting supervised estimation methods. Among supervised methods, in the last few years non-linear regression techniques have obtained a growing attention, thanks to both their good approximation ability even in complex non-linear retrieval problems and their capability to integrate data from different information sources with poorly defined (or unknown) distributions. Examples are machine learning regression methods, like the Multi Layer Perceptron Neural Networks [1]-[3] and the Support Vector Regression [3], [4].

Generally, non-linear regression techniques depend on a vector of free model parameters (or hyper-parameters) $\underline{\theta}$, which have a direct impact on the learning phase of the considered method and thus on the quality of the final estimation. The procedure for tuning these free parameters is often referred as model selection. A commonly adopted and effective strategy to deal with this task consists in selecting, among a set of possible configurations of the free hyper-parameters, the configuration $\underline{\theta}^*$ which optimizes a scalar metric computed on the estimates obtained on a set of reference samples (test or validation samples, which should be independent from the training samples used in the learning phase). The role of the metric is to provide a quantitative and objective evaluation of the quality of the estimation provided by a given model (i.e., configuration of the free hyper-parameters). Possible choices for the metric could be the mean squared error (MSE) or the statistical determination coefficient (R^2) between estimates and reference samples.

However, by definition, different metrics evaluate the quality of the estimation from different perspectives, typically not highly correlated one to each other. Thus, the choice of one specific metric to drive the model selection process implies the lost of the information on the quality of the estimates conveyed by the other metrics. As a consequence, the optimal configuration of the free model parameters selected according to one criterion does not necessarily optimize the other metrics. On the contrary, it is possible that the selected model performs poorly according to other metrics, thus reducing the global quality of the final estimates.

In order to overcome the aforementioned limitations, in this paper we propose to model the issue of tuning the free parameters of a supervised non-linear regression algorithm as a multi-objective optimization problem, in which the multi-objective function is made up of a set of quality metrics. In this framework, multiple solutions to the optimization problem are obtained by jointly optimizing the different considered criteria according to the concept of Pareto optimality. Each solution represents an optimal tradeoff between the different quality metrics. Thus the user has the possibility to effectively individuate and select the model that globally optimizes the quality of the estimates according to the specific requirements of the interested retrieval problem.

2. METHODOLOGY

Let us consider a generic supervised regression algorithm for which a vector $\underline{\theta}=\{\theta_1,\theta_2,...,\theta_n\}$ of n free hyper-parameters should be tuned in order to optimize the quality of the output estimates. $\underline{I}=\{i_1(\underline{\theta}),i_2(\underline{\theta}),...,i_m(\underline{\theta})\}$ is a set of m quality indices or metrics that characterize from different perspectives the quality of the estimates provided by the considered regression technique (e.g., mean squared error, statistical determination coefficient). The simplest strategy to face the model selection issue according to all the m metrics consists in defining a single error function $E(.)$ combining the quality indices through a weighted average:

$$E(\underline{I})=\sum_{j=1}^{m} c_j i_j(\underline{\theta})$$

where $\underline{C}=\{c_1,c_2,...,c_m\}$ is the vector of weights of the average function. The set of hyper-parameters that optimize $E(\underline{I})$ represents the solution to the model selection problem. Despite its simplicity, this formulation has an important drawback: the definition of the weights c_j (which should be done by the user) is very critical because of the intrinsic different scales of the considered metrics. Moreover, the physical information conveyed by the resulting global index is difficult to understand.

In order to overcome these drawbacks, in this work the optimization problem is expressed according to a m-dimensional multi-objective function $\underline{g}(\underline{\theta})$ which is made up of m different objectives $g_1(\underline{\theta}),g_2(\underline{\theta}),...,g_m(\underline{\theta})$ that represent the set of adopted quality metrics computed for different values of the free parameters. All the different objectives of $\underline{g}(\underline{\theta})$ have to be jointly minimized and are considered equally important. Thus the multi-objective optimization problem can be formulated as follows:

$$\min_{\underline{\theta}\in S}\{\underline{g}(\underline{\theta})\}$$

subject to

$$\underline{\theta} = [\theta_1, \theta_2, ..., \theta_n] \in S \subseteq \Re^n$$

This problem is characterized by a vector-valued objective function $\underline{g}(\underline{\theta})$; thus it cannot be solved in order to derive a single solution. On the contrary, a set of optimal solutions $\underline{P}$ can be obtained following the concept of Pareto dominance. More in detail, a generic configuration of the free parameters $\underline{\theta}^*$ is said to be Pareto optimal if it is not dominated by any other configuration in the search space S, i.e., there is no other $\underline{\theta}$ such that $g_i(\underline{\theta}) \leq g_i(\underline{\theta}^*)$ $(\forall i = 1, 2, ..., m)$ and $g_j(\underline{\theta}) < g_j(\underline{\theta}^*)$ for at least one j, $(j = 1, 2, ..., m, j \neq i)$. In other words, $\underline{\theta}^*$ is Pareto optimal if there exists no other subset of free hyper-parameters $\underline{\theta}$ that would decrease an objective without increasing another one at the same time.

Because of the dimensionality and the complexity of the search space, an exhaustive search of the set $\underline{P}$ of optimal solutions is typically unfeasible. As an alternative way, instead of identifying the true set of Pareto optimal solutions, we aim at estimating a set $\underline{P}^*$ of non-dominated solutions with objective values as close as possible to the Pareto front. This estimation can be performed with different multi-objective optimization algorithms that have been proposed in the literature, e.g., multi-objective evolutionary algorithms such as the non-dominated sorting genetic algorithm II (NSGA-II) [5].

The main advantage of the multi-objective approach adopted in this work is that it avoids to aggregate metrics capturing and conveying different and sometimes complementary information on the global quality of the final estimation into a single measure. On the contrary, thanks to the multi-dimensional formulation of the optimization problem, it preserves the physical meaning of each index and allows one to effectively identify different possible optimal tradeoffs between different quality metrics. The final selection of the optimal solution to the model selection problem is demanded to the user, who can identify the best tradeoff among the considered quality metrics on the basis of the specific requirements of the considered retrieval problem.

3. EXPERIMENTAL SETUP AND RESULTS

In order to evaluate the effectiveness on real retrieval problem of the proposed model selection strategy, it was applied to the tuning of the free hyper-parameters of a Support Vector Regression (SVR) estimator for the specific application of soil moisture estimation from microwave remotely sensed data. As quality metrics of the final estimates, different accuracy indices (mean squared error, mean absolute error, etc.), the statistical determination coefficient (R^2) and the shape (slope and intercept) of the linear regression curve between estimates and true measurements were considered. Moreover, in order to investigate the effectiveness and flexibility of the proposed approach to different possible quality requirements of the final user, several experiments were carried out considering different combinations of quality metrics to drive the multi-objective model selection process. An example is reported in Figure 1, which shows a set of estimated Pareto optimal solutions (configurations of the free hyper-parameters of the SVR algorithm) supposing to consider two quality metrics: the Mean Squared Error (MSE) and the determination coefficient (R^2). Each point represents a different tradeoff between the two metrics. For example, configuration 1 provides the best performance in terms of MSE on the test samples, while configuration 2 optimizes the determination coefficient. Solution 3 can be considered as a good compromise between the two quality indices considered: it provides a MSE value similar to configuration 1 but, at the same time, a significant improvement in terms of determination coefficient.

These and other results show that the proposed approach is promising to face the model selection issue for the SVR technique, because: 1) it is based on the information on the quality of the final estimates conveyed by multiple metrics; 2) it allows to effectively derive solutions (configurations of the free hyper-parameters for the regressor) that jointly optimize the metrics selected; 3) thanks to the formulation of the optimization problem, which preserves the physical meaning of each quality index, the user can easily identify, among the available optimal solutions, the one that provides the best global quality of the final estimates according to the specific requirements of the retrieval problem considered.

Further details on the whole experimental analysis will be provided in the full paper.

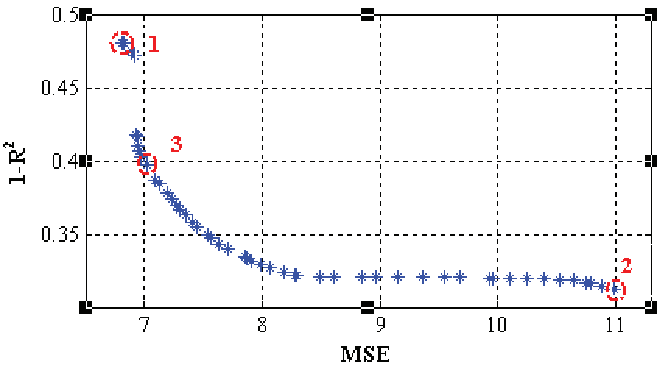

Figure 1. Estimated Pareto optimal solutions considering two quality metrics: MSE and R^2.

4. REFERENCES

[1] C. Notarnicola, M. Angiulli, and F. Posa, "Soil Moisture Retrieval From Remotely Sensed Data: Neural Network Approach Versus Bayesian Method," *IEEE Transaction on Geoscience and Remote Sensing,* vol. 46, no.2, pp. 547-557, February 2008.

[2] N. D. Evora, D. Tapsoba, and D. De Sève, "Combining Artificial Neural Network Models, Geostatistics, and Passive Microwave Data for Snow Water Equivalent Retrieval and Mapping," *IEEE Transaction on Geoscience and Remote Sensing,* vol. 46, no. 7, pp. 1925-1939, July 2008.

[3] L. Pasolli, C. Notarnicola, L. Bruzzone, "Soil Moisture Estimation from Microwave Remote Sensing Data with Non-Linear Machine Learning Techniques," in *Image and Signal Processing for Remote Sensing XV,* edited by Lorenzo Bruzzone, Claudia Notarnicola, Francesco Posa, Proceedings of SPIE Vol. 7477 (SPIE, Bellingham, WA, 2009) 7477 1C.

[4] L. Bruzzone and F. Melgani, "Roboust Multiple Estimator System for the Analysis of Biophysical Parameters From Remotely Sensed Data," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 159-174, January 2005.

[5] K. Deb, A. Pratab, S. Agarwal, and T. Mejarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transaction on Evolutionary Computation,* vol. 6, no. 2, pp. 182-197, April 2002.