ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Adaptive Kernel Smoothing Regression for Spatio-Temporal Environmental Datasets

Federico Montesino Pouzols and Amaury Lendasse *

Aalto University - Dept. of Information and Computer Science
P.O. Box 15400, FI-00076 Aalto, Espoo, Finland

Abstract. This paper describes a method for performing kernel smoothing regression in an incremental, adaptive manner. A simple and fast combination of incremental vector quantization with kernel smoothing regression using adaptive bandwidth is shown to be effective for online modeling of environmental datasets. The method is illustrated on openly available datasets corresponding to the Tropical Atmosphere Ocean array and the Helsinki Commission hydrographic database for the Baltic Sea.

1 Introduction

We describe a method for performing kernel smoothing regression on an adaptive manner. The aim of this work to define efficient, incremental and adaptive regression methods that can be applied sequentially to data streams of incoming observations. The motivation for this work is the need for simple and efficient regression methods that can cope with large, diverse and evolving datasets in applications in Environmental Sciences.

The idea of adaptive regression has been explored in different contexts and a large number of methods for both linear and nonlinear regression are well established in different fields of computer science. For example, the multivariate adaptive regression splines (MARS) method [1, 2] builds models as a summation of weighted basis functions following a divide and conquer strategy that aims to adapt locally. However, most research efforts so far have concentrated on offline regression.

Environmental Sciences have seen a great deal of development and attention over the last few decades. An impressive improvement in observational capabilities and measurement procedures has led to large databases and online monitoring systems. Environmental datasets are normally defined by either regular or irregular spatial fields that can be 3 dimensional, for which multivariate observations such as temperature, salinity, nutrients, pollutants or air pressure, are recorded across time. Environmental processes are usually part of intricate networks of dynamical processes where their evolution in time is a key aspect.

Evolving, online or adaptive intelligent systems [3] are meant to be applied on sequential data or streams of data. These systems distinguish themselves from conventional offline learning methods and previous online methods in that their structure (in addition to their parameters) evolves in order to account for new

*FMP is supported by a Marie Curie Intra-European Fellowship for Career Development (grant agreement PIEF-GA-2009-237450) within the European Community´s Seventh Framework Programme (FP7/2007–2013)..

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

data. Specially during the last decade many advances have been made within the area of evolving neuro-fuzzy systems for modeling and control [3, 4, 5]. Two advantages of these methods are specially relevant for spatio-temporal environmental datasets. On the one hand, they rely on simple and fast algorithms, usually operating in a one-pass manner. On the other hand, their parameters but more importantly their structure evolve in order to accommodate for new data. Thus, large datasets can be efficiently processed online in a fully adaptive manner.

The paper is organized as follows. Section 2 describes the proposed method. The method is evaluated and compared in section 3.

2 Proposed Method

Kernel regression, also called kernel smoothing regression in order to avoid confusion with other kernel methods, is a non-parametric approach to estimating the conditional expectation of a random variable y [2, 6]: $E(y|x) = f(x)$, where y and x are random variables and $f(\cdot)$ is a non-parametric function. The kernel smoothing regression approach is based on kernel density estimation. It is assumed that the model estimation has the following form: $\hat{f}(x) = y + \varepsilon$, i.e., the random variable modeled can be expressed as a deterministic, functional component plus a noise component.

One particular case is the Nadaraya-Watson kernel regression method for function estimation which uses the Gaussian kernel. If n observations of x and y pairs, (x_i, y_i) are available, the estimator of $\hat{f}(\cdot)$ for a particular input observation is defined as follows:

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} K_h(x_0, x_i) y_i}{\sum_{i=1}^{n} K_h(x_0, x_i)},$$

where h is the bandwidth or smoothing parameter, and K_h is the kernel function [7].

Vector Quantization (VQ) is an unsupervised method with parallelisms with methods for clustering and learning densities such as k-means and Voronoi diagrams [2]. It is a practical and popular approach in signal processing and machine learning for lossy data compression and correction but also for density estimation. A key aspect of VQ for the purposes of this work is that it allows to approximate the probability distribution function of a process by the distribution of prototypes or codewords. In fact, the area closer to a particular codeword than to any other is inversely proportional to the density in that region of the input domain.

The approach proposed here is to perform kernel regression on an incremental estimation of the (potentially evolving) probability distribution of the incoming data stream rather than the full sequence of observations. This is done in two stages. First, VQ is incrementally performed on the incoming stream. Second, kernel smoothing regression for each incoming observation is computed using the codebook resulting from the first stage as an estimation of the probability distribution of the incoming data. In addition, the kernel bandwidth is adapted

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

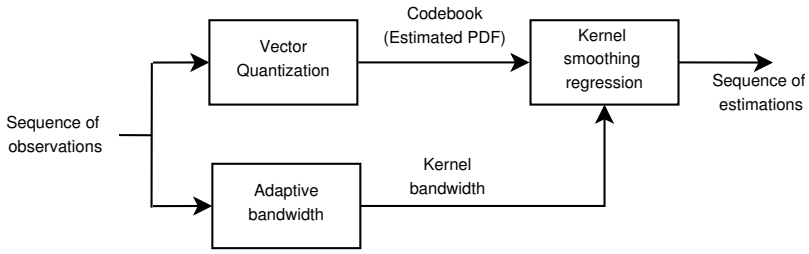


Fig. 1: Global scheme of the KSR-VQ method

online. All the steps required are incremental and the method is thus suitable for online learning. The method is simple, adapts locally to fit evolving streams of data, and is fast, with run-time and memory complexity proportional to the number of observations and their dimensionality. An scheme of this method (KSR-VQ) is shown in figure 1. The two stages of KSR-VQ are detailed in what follows.

2.1 Adaptive Vector Quantization

The first stage of KSR-VQ is performed adaptively and in an incremental manner. Observations are processed one at a time. Let m be the current number of prototypes in the codebook, initialized to 0, and M a maximum number of prototypes. In algorithm 1 we show a simple version of vector quantization which is used in this paper. It should be noted that no sensitivity parameters are used.

Algorithm 1 Simple Vector Quantization

· Initialize an empty codebook.
· For each new observation $\boldsymbol{x}_i, i = 1, \ldots,$ of dimension d received,
with current codebook length m:
IF $m < M$
Add $\boldsymbol{x}_i$ to the codebook as a new prototype.
ELSE
Find in the codebook the nearest prototype, $\boldsymbol{p}_j$, to $\boldsymbol{x}_i$
Update the codebook with learning rate α, move $\boldsymbol{p}_j$ towards $\boldsymbol{x}_i$:
$\boldsymbol{p}_j \leftarrow (1-\alpha)\boldsymbol{p}_j + \alpha\boldsymbol{x}_i, \quad \alpha \in [0,1]$

As it will be shown in section 3, relatively small codebooks of a few hundred prototypes can achieve satisfactory performance in a rather general setup.

2.2 Adaptive Kernel Smoothing Regression

It is generally accepted that local adaptation of the kernel bandwidth parameter is of major importance for obtaining accurate models [6, 7]. However, finding optimal or good values for the bandwidth parameter and furthermore adapting it locally is not a trivial task.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

It is possible to define an estimator for the bandwidth or smoothing parameters optimal for normal distributions [6] in the sense that the mean integrated square error is minimized. For the multivariate case it is as follows:

$$h_{opt,j} = \sigma_j \left(\frac{4}{n(d+2)} \right)^{(1/(d+4))}, \qquad j = 1, \ldots, d,$$

where d is the input dimension, n is the number of observations, and σ_j is the standard deviation in the jth dimension. Nonetheless, as an enhanced standard deviation estimator, the median absolute deviation estimator can be used to approximate σ_j in a robust manner, as described for global bandwidth estimation in [6]. This way, even though the parameter estimators, $h_{opt,j}$, are defined as optimal for normal distributions they still perform well in more general cases.

In KSR-VQ the variance term, σ_j, for each input dimension j is estimated using the median absolute deviation and a scaling factor for normal distributions:

$$\sigma_j = median\{|x_{ij} - median\{x_{ij}\}|\}/0.6745,$$

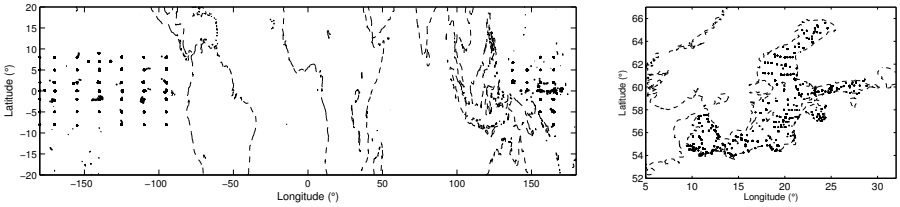
and the median is calculated online for observations $i = 1, \ldots,$ using an incremental statistics method by Manku et al. [8], with approximation guarantees that apply for arbitrary value distributions and arrival distributions of the dataset. This defines a clear criterion for the selection of the bandwidth, i.e., for model selection, without the need for validation procedures. For simplicity, in this paper we restrict our analysis to zero-order or Nadaraya-Watson kernel smoothing regression. That is, KSR-VQ builds nonlinear locally constant models.

3 Experiments and Discussion

Two environmental databases are considered in order to illustrate the suitability of the proposed method for environmental applications, and specially for environmental data streams: UCI El Nino, and the Helsinki Commission (HELCOM) hydrographic database for the Baltic Sea. Figures 2(a) and 2(b) show the location of the observations, respectively. As of October 2010 the HELCOM database consists of 623181 multivariate observations of up to 62 variables, from 1900 up to present time, with implications for multiple fields such as Physical Oceanography, Marine Biology and Climatology [9]. It is available from the oceanographic database of the International Council for the Exploration of the Seas. The UCI El Nino dataset [10] was collected by the Tropical Atmosphere Ocean (TAO) array during the period 1980–1998. The TAO array provides real-time oceanographic and surface meteorological data to scientists, climate researchers and weather prediction centers around the world. This particular dataset corresponds to nearly 70 moored buoys spanning the equatorial Pacific.

In the "UCI El Nino-SST" problem defined from this dataset, the sea surface temperature (SST) has to be modeled as a function of 6 inputs: time, latitude, longitude, zonal and meridional winds, and air temperature. For the HELCOM Baltic database two regression problems are defined corresponding to the dissolved oxygen concentration and salinity of the surface layer (0–25 m).

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.



(a) Buoys of the TAO array. (b) HELCOM cruises.

Fig. 2: Coordinates of measurements, UCI EL Nino and HELCOM Baltic datasets.

Two well known offline methods for regression are included in this study for the purposes of comparison: Multivariate adaptive regression splines (MARS) [1, 2] and multivariate kernel smoothing regression (KSR) [6, 7]. KSR-VQ is also compared against two well known evolving neuro-fuzzy methods: Evolving Takagi-Sugeno (eTS) [3] and DENFIS [5]. First-order Takagi-Sugeno systems are built.

For offline approaches, fitting errors are shown in table 1. The superiority of MARS in terms of accuracy is clear. It comes however at the expense of a longer run-time which can be an order of magnitude higher than that of KSR. Errors are given as normalized root mean square error (NRMSE), i.e., the RMSE divided by the standard deviation of the target sequence, and symmetric mean absolute percentage error (SMAPE). Standard deviations of the absolute errors are indicated as well.

Online regression methods are compared in table 2. Among the conclusions about KSR-VQ that can be drawn from the table are: a) it achieves satisfactory accuracy in general and b) it is consistently the fastest method. The implication of these results is that KSR-VQ would be preferable to standard KSR for offline modeling. In fact, KSR-VQ is competitive against the highly accurate MARS models, but for a lower computational cost in general. Also, it should be noted that the adaptive selection of the kernel parameter or bandwidth in multivariate kernel smoothing regression can be seen as an indirect way of variable scaling.

Table 1: Comparison of offline regression methods.

Dataset	Method	NRMSE	std AE	SMAPE	Run-time (s)
UCI El Nino-SST	MARS linear	4.934e-01	3.387e-01	8.19	3.34e+03
	KSR	5.872e-01	3.981e-01	9.71	1.26e+03
Baltic Dissolved O_2	MARS linear	5.169e-01	3.589e-01	7.79	3.14e+03
	KSR	5.664e-01	4.003e-01	8.38	2.13e+02
Baltic Salinity	MARS linear	2.419e-01	1.887e-01	10.8	3.00e+03
	KSR	4.045e-01	3.023e-01	16.8	1.80e+02

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Table 2: Comparison of adaptive regression methods.

Dataset	Method	NRMSE	std AE	SMAPE	Run-time (s)
UCI El Nino-SST	DENFIS	**2.388e-01**	1.832e-01	3.50	1.42e+03
	eTS	6.287e-01	4.309e-01	10.2	5.54e+03
	KSR-VQ	3.107e-01	9.653e-02	4.21	1.59e+02
Baltic Dissolved O_2	DENFIS	4.349e-01	3.364e-01	5.94	3.95e+02
	eTS	6.470e-01	4.400e-01	9.85	8.72e+02
	KSR-VQ	**3.832e-01**	2.816e-01	5.40	2.58e+01
Baltic Salinity	DENFIS	2.915e-01	2.335e-01	11.3	5.65e+02
	eTS	4.300e-01	2.928e-01	24.7	8.86e+02
	KSR-VQ	**2.802e-01**	2.441e-01	8.46	2.48e+01

4 Conclusion

The adaptiveness of KSR-VQ is twofold. First, the smoothing regression is performed on an incrementally updated, evolving estimation of the probability distribution of the incoming data stream. Second, the kernel bandwidth is adapted online using a criterion based on the median absolute deviation estimator which can be computed efficiently online. The advantages of KSR-VQ against offline KSR as well as its competitiveness with established adaptive and evolving methods have been illustrated for two environmental datasets. KSR-VQ shows good generalization capabilities in spite of its simplicity and low computational requirements.

References

[1] J. Friedman. Multivariate adaptive regression splines. *Ann. of Stat.*, 19(1):1–141, 1991.

[2] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, NY, USA, 2nd edition, 2009.

[3] P. Angelov, N. Kasabov, and D. Filev, editors. *Evolving Intelligent Systems: Methodology and Applications*. Wiley-IEEE Press, 2010. IEEE Series on Computational Intelligence.

[4] F. M. Pouzols and A. Lendasse. Evolving fuzzy optimally pruned extreme learning machine for regression problems. *Evolving Systems*, 1(1):43–58, 2010.

[5] N. Kasabov and Q. Song. DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction. *IEEE Trans. Fuzzy Syst.*, 10(2):144–154, 2002.

[6] A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis: the kernel approach*. Oxford University Press, 1997.

[7] J. S. Simonoff. *Smoothing Methods in Statistics*. Statistics. Springer, 1998.

[8] G. S. Manku, S. Rajagopalan, and B. G. Lindsay. Approximate medians and other quantiles in one pass and with limited memory. *ACM SIGMOD Record*, 27(2), 1998.

[9] The BACC Author Team. *Assessment of Climate Change for the Baltic Sea Basin*. Regional Climate Studies. Springer-Verlag, 1st edition, 2008.

[10] A. Frank and A. Asuncion. UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/, 2010. University of California, Irvine, School of Information and Computer Sciences.