Improving Laplacian Pyramids Regression with Localization in Frequency and Time

Ben Hen[1], Ángela Fernández[2] and Neta Rabin[1]

1 - Department of Industrial Engineering, Tel-Aviv University, Israel.
2 - Dpto de Ingeniería Informática, Universidad Autónoma de Madrid, Spain.

Abstract. Auto-Adaptive Laplacian Pyramids (ALP) is an iterative kernel-based regression model. It constructs a multi-scale representation of the train data, where the multi-scale modes are average residuals. In this work, we propose two extensions of the model. The first is a hybrid approach that combines ALP with Empirical Mode Decomposition to provide localization in the frequency domain. The second modifies ALP to fit datasets with non-uniform noise, which is achieved by computing the optimal stopping criterion in a point-dependent manner. Experimental results demonstrate these models for solar energy prediction and for forecasting epidemiology infections.

1 Introduction

Kernel based regression is a powerful forecasting tool that is applied in varied domains like the energy market and epidemiology forecasting. The Laplacian pyramids algorithm, which was originally introduced for image coding by Burt and Adelson [1], may be easily utilized to work as a multi-scale regression model [2]. It convolves the data with Gaussian kernels of decreasing widths to yield a multi-scale representation. In statistics, this approach is known as the Nadaraya-Watson estimator [3]. The stopping scale can be determined by using L_2 boosting, as described in [4, 5] ,and we denote this scheme by Auto-Adaptive Laplacian Pyramids (ALP).

In this work, we focus on two improvements of the ALP model. The first takes a hybrid approach by combining ALP with the Empirical Mode Decomposition (EMD) [6]. The second reviews and demonstrates the ALP-local model that was recently introduced in [7]. The EMD-ALP hybrid model enhances performance by localizing in frequency and the ALP-local model is constructed based on local, rather than global information in the spatial (or time) domain. These extensions are suited to address different types of challenges in time series forecasting tasks, like dealing with noise that is distributed non-uniformly or present only in high-frequencies of the data.

The hybrid EMD-ALP model first evokes EMD. EMD decomposes time series into intrinsic modes, named IMFs, and a residual. Its main advantage is its ability to handle nonlinear and non-stationary time series, as well as the simplicity of its construction. Over the last years, several hybrid models that combine EMD with known forecasting techniques were proposed. In [8], a hybrid EMD and Kernel-Ridge Regression (KRR) model was used for wind power prediction, and in [12] EMD was combined with SVR for energy load forecasting. EMD was

also shown to improve the performance of an ARIMA - Artificial Neural Network (ANN) model in [9]. In this work we demonstrate the EMD-ALP model for the prediction of solar energy. We compare our results to other hybrid EMD based methods like EMD-KRR (Kernel Ridge Regression) and EMD-SVR (Support Vector Regression) and highlight the performance of ALP. Since the iterative mechanism of ALP may be used as an interpolating or as a regression model [10], the application of ALP on the low-pass IMF modes, yields an approximation model that behaves close to an interpolation scheme, resulting with low forecasting errors. Moreover, the Gaussian kernels extend the average regression residuals in each scale, thus producing a relatively smooth prediction function.

The second challenge we address is modeling of time series with a non-uniform noise distribution. Examples of such time series come from the field of epidemiology, where it is desired to predict the number of future infected individuals. When the disease reaches a *wave*, the numbers may fluctuate daily, resulting in a noise time series. Once the *wave* is over, we observe a period with a very small number of new infections, hence the time series is relatively smooth. The ALP scheme that was proposed in [2, 5] stops the refinements based on the global L_2 error, thus each data point is decomposed into an equal number of modes. In [7], a modified local stopping scale was introduced. It selects the optimal stopping scale for each point, with respect to the local noise level and the density. In this work, we utilize this approach for forecasting the number of future infected individuals from chickenpox in the city of Budapest, Hungary.

The rest of the paper is organized as follows. In Section 2 we describe the ALP, ALP-local and the EMD methods. Section 3 presents an application for solar energy prediction using the EMD-ALP hybrid model. In Section 4 prediction of future chickenpox infection cases using the ALP-local model is described. Conclusions are provided in Section 5.

2 Mathematical Background

ALP is a general method for function approximation and extension. Here, we describe its setting in the way it was used in the experimental results. Given a time series data $y(t)$, where $1 \leq t \leq n$, we wish to make a future prediction, $y(n+1)$ based on short-term trajectories from $y(t)$. Denote the set of overlapping short-term trajectories of length k by $X = \{x(t,:)\}_{t=1}^{n-k+1}$, that were constructed using an overlapping sliding window over $y(t)$. The training set is composed of pairs $\{x(t,:), f(t)\}$, where $f(t) = y(t+k+1)$ is the target. Given a new sample $x(\tilde{t},:)$, the task is to predict $f(\tilde{t})$. For simplicity, we denote $x(t,:)$ by x_t.

2.1 Auto-adaptive Laplacian Pyramids

ALP produces a multi-scale representation of the function f, based on modified Gaussian kernels that are constructed on the pairwise distances of X. A series of Gaussian kernels of decreasing scales, $G_l = (g_l(x_t, x_{t^*}))$, are defined by

$$G_l = g_l(x_t, x_{t^*}) = e^{\frac{-\|x_t - x_{t^*}\|^2}{\sigma_l}}, \tag{1}$$

where $x_t, x_{t^*} \in X$. The initial kernel scale σ_0 is set to be large, for example, $\sigma_0 = C \cdot \max_t \left[\min_{t^*, t^* \neq t} \left(\|x_t - x_{t^*}\|^2 \right) \right]$, and $\sigma_l = \frac{\sigma_{l-1}}{2}$. In order to incorporate the L_2 boosting into the model and to automatically detect an appropriate global stopping scale for the model, the kernels are modified as follows:

$$\tilde{g}_l(x_t, x_{t^*}) = \begin{cases} g_l(x_t, x_{t^*}) & t \neq t^* \\ 0 & t = t^*. \end{cases}$$

We denote the associated row-normalized kernel of $\tilde{G}_l$ by K_l. Then, f is smoothed in an iterative manner using the series of smoothing operators K_l and a course representation is constructed by $s_0(x_t) = \sum_{t^* \in X} k_0(x_t, x_{t^*}) f(x_t)$. The difference $d_1 = f - s_0$ is averaged by a finer kernel, resulting with s_1, and a finer representation of f is $f_1 = s_0 + s_1$.

In general, for $l = 1, 2, 3 \ldots$, we have $d_l = f - f_{l-1}$, and $f_l = f_{l-1} + s_l$. The stopping scale is set by computing the mean square error at each level and choosing the scale $l^\star$ for which the minimum error value occurs.

2.2 The ALP-local model

The stopping scale in the ALP model is using the mean square error at each level, denoted by $\mathrm{err}^{(l)}$, and results with one global stopping scale, $l^\star$, that is associated with all of the data points. In [7], both the model construction and the extension procedure were modified to provide the freedom for assigning a point-wise stopping scale. A new parameter ν was added to the model. It defines the size of the local neighborhood in X upon which the point-wise error is computed. In other words, instead of computing a global error value, $\mathrm{err}_t^{(l)}$, at each iteration, a point-wise local error $\mathrm{err}_t^{(l)}$ is computed at each scale by considering the point's nearest neighbors. The vector that holds the final point-wise stopping scales is defined by $l_t^\star = \arg\min_l \{\mathrm{err}_t^{(l)}\}$. When forecasting the function value for a new time-trajectory point $x_{\tilde{t}}$, we identify its nearest neighbor in X and set $l_{\tilde{t}}^\star$ to be the stopping scale that is associated with its nearest neighbor.

2.3 Empirical Mode Decomposition

EMD models the time series $x(t)$ as a linear and finite combination of components termed Intrinsic Mode Functions (IMF). There are no a priori assumptions regarding the nature of the data. For a given time series $x(t)$, EMD finds a set of L IMFs $\{C_i(t)\}_{i=1}^L$ and a residual signal $r(t)$, so that

$$x(t) = \sum_{i=1}^{L} C_i(t) + r(t). \tag{2}$$

The IMFs are computed by the following *sifting* process. First, all of the local extremes of $x(t)$ are identified. Then, two envelopes are created from the minima and maxima extreme points by application of the Cubic Spline. The

envelopes, denoted by $e_{\min}(t)$ and $e_{\max}(t)$ cover all the data between them. The mean series of the envelopes, $m(t)$, is computed, followed by a computation of $d_1(t) = x(t) - m(t)$. The process is then iterated over $d_i(t)$ until a stopping criteria determines if $d_i(t)$ is a legal IMF. If the criteria is fulfilled then $d_i(t)$ is set to be the first IMF. A residual defined by $r(t) = x(t) - c(t)$ is computed and the entire process is repeated. The process stops when the residual becomes a monotonic function, and this last residual is set to $r(t)$.

3 An EMD-ALP Model for Prediction of Solar Energy

Our hybrid model applies the ALP model over the EMD decomposition of the time series data. In this process, the training set X is kept fixed, while the function f is replaced with its EMD decomposition (see Eq. (2)). Traditionally, predictions are made for each mode separately, including the residual, and then summed up. Since we did not see a clear improvement with this straight-forward combination, we suggest an alternative hybrid version. We calculated the average standard deviation of overlapping windows in each IMF when considering window lengths of size 5 and 15. For each window configuration, we split the IMFs (including the residual) into two groups that share similar variability. The IMFs in each group were aggregated, yielding two intrinsic modes. One holds higher frequencies and the other holds lower frequencies. We used a validation set (taken from the training samples) for selecting a grouping that enhanced the prediction accuracy. Note that the default EMD configuration we used generates approximately 7 IMFs that can be fed as input to the regression models, while our approach feeds the model with two aggregated IMFs. Denote this EMD model by EMD_A.

The EMD-ALP approach is demonstrated on the AMS 2013-2014 Solar Energy Prediction Contest (Kaggle). The goal was to predict the total daily incoming solar energy. The data represents the daily aggregated radiation from 98 stations in Oklahoma between 1994-2007. For this experiment, we selected five batches of size 98×600, which were converted into an overlapping time series of length 7, for each station. The models were created based on the first 300 time trajectories, where a 200-100 split was used for setting the window length size hyperparamter. The remaining 300 trajectories were test points. We compared our model against Kernel Ridge Regression (KRR) with an RBF kernel, EMD-KRR, Support Vector Regression (SVR) and EMD-SVR. EMD-KRR and EMD-SVR were worse than the KRR and SVR models, thus we omitted these ones from the results table. Table 1 presents the average results for the 5 batches in terms of root mean square error. It can be seen that ALP produces accurate results that are further improved with our EMD_A decomposition.

4 ALP-local for Predicting Chickenpox Cases

The chickenpox dataset [11] includes weekly chickenpox cases from 20 cities in Hungary. We performed predictions for Budapest. The series was decomposed

KRR	SVR	ALP	EMD-ALP	EMD$_A$-ALP
5363311.7	8082705.9	3066830.9	3115669.07	2775752.2

Table 1: Prediction Errors (RMSE) for the Solar Dataset

	KNN	SVR	ALP	ALP-local
RMSE	75.2096	69.3583	63.6401	63.2275
MAE	48.0600	42.9782	43.4024	41.5525

Table 2: Prediction Errors for Budapest Chickenpox

into overlapping windows of length 5. The prediction of each test point was done via building a model on the previous 100 train trajectories. ALP-local was executed with $\nu = 25$, after performing a grid search with $\nu \in \{15, 25, 35, 45\}$ with a different part of the data.

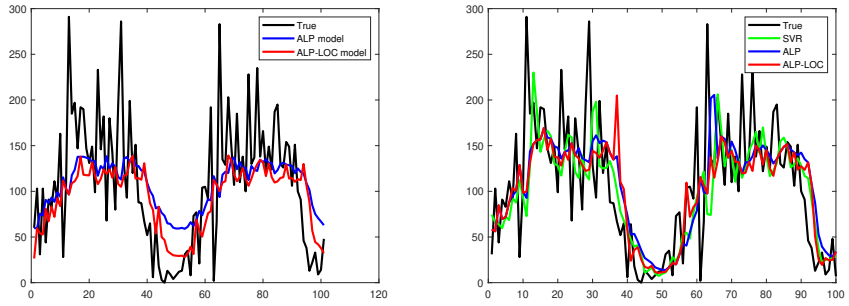

Fig. 1: Left: ALP (blue) and ALP-local (red) train models, Right: ALP (blue), ALP-Local (red) and SVR (green) predictions.

Figure 1 (left) shows the 100 train function samples $f(t)$ from Budapest. The ALP train model is plotted in blue and the ALP-local train model is plotted in red. It can be seen that ALP does not reach the smooth areas of $f(t)$, this is due to the high noise that affects the global stopping scale. Here ALP stopped after 6 iterations while ALP-local model contained a maximal scale of 7 for points that belong to the smooth regions, and scales between 5-6 for the noisy parts. Figure 1 (right) presents the predictions for 100 test points. In addition to the two ALP models, we plot the SVR predictions in green and it can be seen that it produces some spikes and that it predicts the smoother areas as well as ALP-local. The numerical results for 100 test points, in terms of RMSE and MAE(L_1 norm), are shown in Table 2, where a KNN model ($K = 5$) has been also used. It can be seen that ALP-local performs better than the other compared methods.

5 Conclusions

We proposed two extensions for the ALP model that offer some advantages in different contexts. The first EMD-ALP model uses the locality in the frequency domain that is achieved by EMD for constructing more accurate ALP models. Furthermore, we showed that the application of ALP on the aggregated EMDs, which are grouped in a way that optimizes the training phase, is favorable. The second extension re-visited the recently introduced ALP-local model and demonstrated its advantage for the prediction of epidemiological time series, a dataset with non-uniform noise.

Acknowledgements

This research was supported by the Israel Science Foundation [Grant 1144/20].

References

[1] P. Burt and E. Adelson, The Laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[2] N. Rabin and R. R. Coifman, Heterogeneous datasets representation and learning using diffusion maps and Laplacian pyramids, *Proceedings of the 2012 SIAM International Conference on Data Mining*, 189–199, 2012.

[3] H. J. Bierens, The Nadaraya-Watson kernel regression function estimator, *Topics in Advanced Econometrics* New York: Cambridge University Press. 212–247, 1994.

[4] M. Di Marzio and C. C. Taylor, On boosting kernel regression, *Journal of Statistical Planning and Inference*, 138(8):2483–2498, 2008.

[5] Á. Fernández, N. Rabin, D. Fishelov and J. Dorronsoro, Auto-adaptive Laplacian pyramids, *Proceedings of the 24th European Symposium on Artificial Neural Networks*, Bruges, Belgium, 59–64, 2016.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N-C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995, 1998.

[7] Á. Fernández, N. Rabin, D. Fishelov and J. Dorronsoro, Auto-adaptive multi-scale Laplacian Pyramids for modeling non-uniform data, *Engineering Applications of Artificial Intelligence*, 93:103682, 2020.

[8] J. Naik, P. Satapathy and P. K. Dash, Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression, *Applied Soft Computing*, 70:1167–1188, 2018.

[9] Ü. Ç. Büyükşahin and Ş. Ertekin, Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition, *Neurocomputing*,361:151–163, 2019.

[10] L. Kang and R. V. Joseph, Kernel approximation: From regression to interpolation, *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):112–129, 2016.

[11] B. Rozemberczki, P. Scherer, O. Kiss, R.Sarkar and T. Ferenci, Chickenpox Cases in Hungary: a Benchmark Dataset for Spatiotemporal Signal Processing with Graph Neural Networks, *arXiv preprint arXiv:2102.08100*, 2021.

[12] L. Ghelardoni, A. Ghio and D. Anguita, Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression, *IEEE Transactions on Smart Grid*, 4(1):549–556, 2013.