ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Multi-Directional Laplacian Pyramids for Completion of Missing Data Entries

Neta Rabin

Department of Industrial Engineering, Tel-Aviv University, Israel.

Abstract. A common pre-processing task in machine learning is handling missing data entries, also known as imputation. Standard techniques use mean values, regression or optimization based techniques for predicting the missing data values. In this paper, a kernel based technique is utilized for imputing data in a multi-scale manner. The construction is based on Laplacian pyramids, which operate on the row and column spaces of the data in several scales. Experimental results demonstrate the approach on publicly available datasets, and highlight its simple computational construction and convergence stability.

1 Introduction

Kernel-based data modeling has become a standard way for coding the relationship between data instances that originate from real-life phenomena. Kernels play a central role in many unsupervised algorithms aiming for finding a compact representation, which reflects the underlying phenomena of the dataset. Non-linear dimensionality reduction methods such as Eigenmaps [1], Diffusion Maps [2], and Magnetic Eigenmaps [3], evokes kernels which capture the local geometric structure of the data. The kernels' spectral decomposition is then computed for deriving a set of reduced coordinates. Once the desired compact data representation is achieved, it may be utilized for evaluating and predicting functions that are defined on the data. Regression methods provide a simple mechanism for evaluating empirical functions over scattered data points. In particular, kernel based regressions [4, 5] are suitable for cases in which the relationship between the data points and the function is not linear.

This paper is focused on a multi-scale kernel based technique for evaluating functions over scattered data. The construction is based on Laplacian pyramids (LP), which were originally introduced for image coding [6]. Modeling scattered datasets with LP was proposed in [7]. The method builds a multi-scale data representation, which can be extended to handle newly arrived data points at each scale in an efficient and a stable way.

Auto-adaptive Laplacian pyramids (ALP), proposed in [8], improved the original LP method. The ALP scheme prevents over-fitting and is robust to noise by automatically detecting a data-adapted optimal stopping scale for the model. The method was demonstrated on a noisy time series data set for predicting solar energy. ALP together with diffusion maps were proposed for completion of missing data in high-dimensional datasets [9]. A new two-sided ALP method was proposed in [10] for learning multi-scale connectivity patterns in a matrix shaped datasets.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Building on the described LP framework, this work proposes a modified, flexible, multi-scale technique for multi-directional analysis of high-dimensional datasets. The concrete application is missing data completion (or imputation), which is a common pre-processing task in machine learning. Straightforward approaches include mean and median based imputations [11]. Imputation is also related to the matrix completion problem, which may be addressed by using nuclear norm minimization [15], or by generalizations to other regularization [16]. An unsupervised manifold learning method that reveals the underlying geometry of a given matrix based on multi-scale information was presented in [17]. This method can be also applied in case of missing data. Here, a new direction for multi-scale imputation is proposed. The method does not suffer of convergence issues, it is easy to implement and it may handle a large number of missing entries.

This paper is organized as follows. In Section 2 the LP and ALP models are described. In Section 3 the new multi-directional approach is outlined. The adaptation to data imputation together with experimental results are presented in Section 4. Conclusions and future extensions are provided in Section 5.

2 Mathematical Background

2.1 One-directional Laplacian Pyramids

Let $X = \{x_1, \ldots, x_n\}$ be a set of scattered data points, possibly high dimensional, and let f be a function defined on X. The LP representation of f is an iterative construction, which is defined as follows. A coarse Gaussian kernel $G_0 = (g_0(x_i, x_j))$, having a large scale of σ_0, defined by $g_0(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{\sigma_0}}$, $x_i, x_j \in X$, is constructed from X. The associated row-normalized kernel of G_0 is given by $K_0 = (k_0(x_i, x_j))$, where $k_0(x_i, x_j) = \frac{g_0(x_i, x_j)}{\sum_k g_0(x_i, x_k)}$.

At a finer scale l, the kernel G_l is defined by $g_l(x_i, x_j) = e^{-\|(x_i - x_j)\|^2/(\frac{\sigma_0}{2^l})}$. Normalization of G_l yields the smoothing operator K_l. The first iteration, for which $l = 0$, generates a smooth approximation of f that is given by

$$f_0(x_i) = \sum_{j=1}^{n} k_0(x_i, x_j) f(x_j), \quad i = 1, \ldots, n, \quad x_i, x_j \in X. \tag{1}$$

In order to simplify the mathematical formulation, the discrete convolution sums will be written using continuous notations, resulting in a continues form for Eq. (1) written as $f_0 = K_0 * f$.

Let $d_1 = f - f_0$, then a finer representation of f is $f_1 = f_0 + K_1 * d_1$. In general, for $l = 1, 2, 3 \ldots$, we have $d_l = f - f_{l-1}$, and $f_l = f_{l-1} + K_l * d_l$, where f_0 is defined in Equation (1). The function f is approximated by the series of functions $\{f_0, f_1, f_2, \ldots\}$ in a multi-scale manner, going from a coarser to a finer representation. The functions $\{f_0, f_1, f_2, \ldots\}$ can be easily extended to a new point $\bar{x}$ by evaluating the multi-scale kernels for the pairwise distances between X and $\bar{x}$ and then convolving with the known function values.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

2.2 Two-directional Laplacian Pyramids

Given a matrix type dataset X of size $M \times N$ and function $f = f(x, y)$ of size $M \times N$ that is defined on X, the LP method is modified to a two-directional approximation [10]. At each scale l, two normalized kernels are constructed, these are denoted by $K_l^{(L)}$ and $K_l^{(R)}$, for left (L) and right (R) respectively. First, f is coarsely approximated by $f_0 = K_0^{(L)} * f * K_0^{(R)}$. Next, the difference $d_1 = f - f_0$, is calculated and it becomes the input for the next finer approximation $f_1 = f_0 + K_1^{(L)} * d_1 * K_1^{(R)}$. After $l - 1$ iterations the difference between f and its multi-scale representation is given by $d_l = f - f_{l-1}$, and a fine version of f is $f_l = f_{l-1} + K_l^{(L)} * d_l * K_l^{(R)}$.

2.3 Auto-adaptive Laplacian Pyramids

The LP iterations may be stopped by setting an admissible error to a small threshold, defined by *err*, and requiring $\|f - f_l\| < err$. A modification of this procedure (see [8]) determines an appropriate stopping scale by implementing a leave-one-out-cross-validation within the algorithm. This modification treats each training point $x \in X$ is treated as a test point, which prevents over-fitting. The multi-scale kernels are modified by setting a zero-diagonal, defined by

$$\tilde{G}_l(x_i, x_j) = \begin{cases} G_l(x_i, x_j) & i \neq j \\ 0 & i = j. \end{cases} \tag{2}$$

The corresponding row-normalized operators are denoted by $\tilde{K}_l$ for the one-directional scheme and $\tilde{K}_l^R$, $\tilde{K}_l^L$ for the two-directional setting. 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.

2.4 Laplacian Pyramids Error Analysis

Theorems regarding the error rate of LP were proved in [10]. It was shown that if a function f is in $W^{2l^\star,2}$, (the space for which f and its weak derivatives are in L_2), where $l^\star$ is the final stopping level, then the LP error obtains

$$\|d_l(x)\|_{L^2} \leq C\sigma_0^2 \left(\frac{\sigma_0^2}{\mu^l}\right)^{l-1} \|f(x)\|_{2l,2}\,.$$

Recently, in [18] conditions for the algorithm's convergence were provided together with stability bounds on the extended function.

3 Multi-Directional Laplacian Pyramids

Given a dataset X of size $M \times N$ and a function $f = f(x, y)$, which is defined on X, we propose the following multi-directional LP construction. For level l_0, a coarse approximation of f, given by

$$f_0 = \frac{1}{2}\left(\tilde{K}_0^{(L)} * f + f * \tilde{K}_0^{(R)}\right),$$

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

is constructed. The normalized zeros-diagonal kernels $\tilde{K}_0^{(L)}$ and $\tilde{K}_0^{(R)}$ capture the pairwise distances between the rows and the columns of X, respectively. Then, the difference $d_1 = f - f_0$ is input for the next iteration, which yields $f_1 = f_0 + \frac{1}{2}\left(\tilde{K}_1^{(L)} * d_1 + d_1 * \tilde{K}_1^{(R)}\right)$. After l iterations, the functions f is expressed by

$$f_l = f_{l-1} + \frac{1}{2}\left(\tilde{K}_l^{(L)} * d_l + d_l * \tilde{K}_l^{(R)}\right).$$

Furthermore, we present a more general setting, in which one can control the contribution of the row and column kernels to the sum. In this setting f_l is expresses by

$$f_l = f_{l-1} + \left(\alpha\left(\tilde{K}_l^{(L)} * d_l\right) + (1-\alpha)\left(d_l * \tilde{K}_l^{(R)}\right)\right), \tag{3}$$

where $0 \leq \alpha \leq 1$. In this work, α is set by testing several optional values.

4 Completion of Missing Data Entries

The proposed approach is demonstrated for imputation in matrix type datasets. In this setting the function is just the data itself, i.e $f = X$. The data $X_{M\times N}$ includes missing values with known locations, coded in a binary matrix $B_{M\times N}$, where $(b_{ij}) = 1$ indicates known data locations and 0 otherwise. The kernels' pairwise distances are computed only based on the known values, i.e. $(b_{ij}) = 1$.

Application of multi-directional LP as described in Eq. (3) to public datasets from the UCI repository [19] is demonstrated. The features (columns) of the inspected dataset were normalized to have mean 0 and standard deviation 1. Results plot the average error from 10 executions. Missing data locations were chosen at random in each iteration.

Three small datasets are considered: Ecoli (336×7), White wine (4898×11) and Boston housing (506×11, after eliminating two categorical attributes). The results are compared with one-directional LP (1D Pyds), two-directional LP (2D Pyds), and with results from [14]. The proposed multi-directional method is denoted by MD Pyds. The MD Pyds method, which implements Eq. (3), was tested with 3 different α values, $\alpha \in \{0.2, 0.5, 0.8\}$. Best results were achieved for $\alpha = 0.8$. Since the evaluated data sets have a small number of columns, it is reasonable that the model which emphasises the row-convolution kernels yields the best results. The last three columns in Table 1 report the errors for the Iterative Step-wise Regression Imputation method (IRMI) [13], Optimized Linear Imputation (OLI) [14], and the Multiple Imputation technique (MICE) [12]. The missing data percentage in these examples is 5%. IRMI didn't converge for the white wine dataset. The proposed multi-directional LP approach performs well, and in two cases achieves lower errors compared to the other regression based methods.

A lager dataset, voice rehabilitation [19] of size 126×309 in considered. 126 is the number of patients and 309 is the number of computed features. Three

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Dataset	MD Pyds	2D Pyds	1D Pyds	IRMI	OLI	MICE
Ecoli	**0.55**	0.62	0.66	8.26	5.75	1.2
Wine	**0.40**	0.43	0.46	–	0.87	1.1
Housing	0.29	0.32	0.34	**0.28**	0.30	0.56

Table 1: MSE Imputation errors for the three public dataset

experiments were performed, with 20% 50% and 80% of randomly selected missing data entries. The multi-directional model was tested with $\alpha \in \{0.2, 0.5, 0.8\}$ (see Eq. (3)). The described results are for $\alpha = 0.5$, which yields the lowest error rates. The results presented in Table 2 were compared with one-directional and two-directional LP and with imputation by using the mean column value and with the the column's most frequent value. Figure 1 plots the multi-scale process for 300 test points at scales $l = 2, 4$ and 6, where $l^\star = 6$ is the finest scale.

% missing	MD Pyds	1D Pyds	2D Pyds	Mean	Freq.
20% (MSE)	**0.3301**	0.5843	0.5793	0.9954	10.9006
50% (MSE)	**0.4363**	0.7047	0.6788	1.0206	8.2944
80% (MSE)	**0.6093**	0.8408	0.8103	1.0433	5.0751

Table 2: MSE Imputation errors for the voice dataset

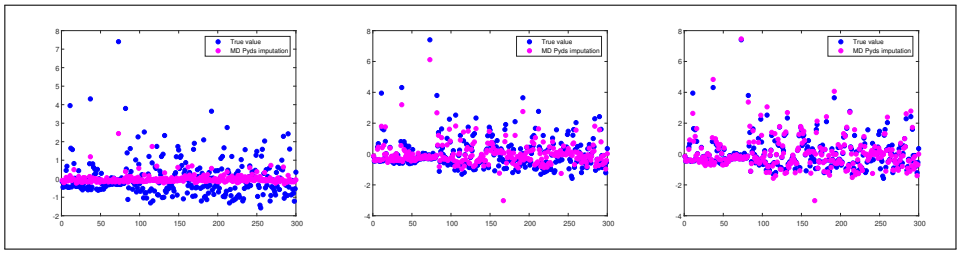

Fig. 1: MD Pyds imputation of 300 test points in iterations 2,4, and 6 from left to right. Known values are colored blue, imputed values are pink.

5 Conclusions

The proposed formulation of the LP approach separates the row and column multi-scale operators, which results in a more flexible scheme. The method is adapted to the task of completions of missing data entries, were it is desirable to take into account the geometric structure of both the row and column spaces. The scheme does not suffer of convergence risks and it significantly reduces the

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

errors when compared to the previous LP constructions. Additionally, it may be easily extended to process cubic or even higher-dimensional datasets with missing entries. In future work, we plan to develop a data-driven criteria for setting the value of the parameter α in Eq. (3), while expanding the experimental results.

References

[1] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Computation*, 15:1373–1396, 2003.

[2] R. R. Coifman and S. Lafon, Diffusion maps, *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

[3] M. Fanuel, C. M. Alaíz, Á. Fernández and J.A.K. Suykens, Magnetic eigenmaps for the visualization of directed networks, *Applied and Computational Harmonic Analysis*, 44(1):189–199, 2018.

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

[5] A. B. Lee and R. Izbicki, A spectral series approach to high-dimensional nonparametric regression, *Electronic Journal of Statistics*, 10(11):423-463, 2016.

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

[7] 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.

[8] 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.

[9] N. Rabin and D. Fishelov, Missing data completion using diffusion maps and Laplacian pyramids, *Computational Science and Its Applications (ICCSA)*, Lecture Notes in Computer Science, vol. 10404, Springer, 284-297, 2017.

[10] N. Rabin and D. Fishelov, Two directional Laplacian pyramids with application to data imputation, *Advances in Computational Mathematics*, 45(4):2123–2146, 2019.

[11] J. A. R. Little and B. D. Rubin, Statistical analysis with missing data, *2nd Edition Wiley*, 2002.

[12] S. Buuren and K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in r, *Journal of statistical software*, 45(3):1–67, 2011.

[13] M. Templ, A. Kowarik and P. Filzmoser, Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, 55(10):2793–2806, 2011.

[14] Y. S. Resheff and D. Weinshall, D., Optimized linear imputation, *Proceedings: 6th International Conference on Pattern Recognition Application and Methods (ICPRAM)*, Porto Portugal, 2017.

[15] E. J. Candes and T. Tao, The power of convex relaxation: Near-Optimal Matrix Completion, *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[16] G. Shabat, Y. Shmueli and A. Averbuch, Missing entries matrix approximation and completion, *Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA)*, 440-443, 2013.

[17] G. Mishne, E. C. Chi and R. R. Coifman, Co-manifold learning with missing data. *Proceedings of the* 36*th International Conference on Machine Learning*, Long Beach, California, 2019.

[18] W. Leeb, Properties of Laplacian pyramids for extension and denoising, *arXiv preprint arXiv:1909.07974*, 2019.

[19] http://archive.ics.uci.edu/ml/datasets.