ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0.
Available from http://www.i6doc.com/en/.

Sparse minimal learning machine using a diversity measure minimization

Madson L. D. Dias[1], Lucas S. Sousa[2], Ajalmar R. da Rocha Neto[2],
César L. C. Mattos[1], João P. P. Gomes[1] and T. Kärkkäinen[3] *

1. Federal University of Ceará
Department of Computer Science, Brazil
{madson.dias, jpaulo}@lia.ufc.br, cesarlincoln@dc.ufc.br

2. Federal Institute of Ceará
Department of Teleinformatics, Brazil
{lucas.sousa, ajalmar}@ppgcc.ifce.edu.br

3. University of Jyvaskyla,
Faculty of Information Technology, Finland
tommi.karkkainen@jyu.fi

Abstract.
The minimal learning machine (MLM) training procedure consists in solving a linear system with multiple measurement vectors (MMV) created between the geometric configurations of points in the input and output spaces. Such geometric configurations are built upon two matrices created using subsets of input and output points, named reference points (RPs). The present paper considers an extension of the focal underdetermined system solver (FOCUSS) for MMV linear systems problems with additive noise, named regularized MMV FOCUSS (regularized M-FOCUSS), and evaluates it in the task of selecting input reference points for regression settings. Experiments were carried out using UCI datasets, where the proposal was able to produce sparser models and achieve competitive performance when compared to the regular strategy of selecting MLM input RPs.

1 Introduction

In the last years, a new supervised learning algorithm, the minimal learning machine (MLM, [1, 2]), has gained attention due to its simple and easy implementation and because its requiriment of only one hyperparameter.

The MLM learning algorithm can be decomposed into two main steps: *distance regression* and *output prediction*. The distance regression consists of solving a linear inverse problem with multiple measurement vectors (MMV), created between the geometric configurations of points in the input and output spaces. Such geometric configurations are built upon through two matrices, created using subsets of input and output points, called reference points (RPs). Output prediction for a new incoming input is achieved by estimating distances in output space using the underlying linear model followed by a search/optimization

*The authors would like to thank UFC, FUNCAP, and Academy of Finland (grants 311877 and 315550) for supporting their research.

procedure in the space of possible outputs. This problem can be understood as a multilateration problem [3] to estimate the new point location in the output space using the approximated distances.

Previous works shows that the determination of the RPs, including their quantity, is fundamental to the generalization of the MLM model for classification tasks [4, 5]. In this regard, the training algorithm for the original MLM establishes that the input and output RPs are associated to the same data samples and their selection is made randomly, leaving just the number of points as a user-selected parameter [1]. Although this selection method can reach good numerical results, the random selection and the use the same pairs of reference points in the input and the output spaces may be a sub-optimal choice, since the RPs have different purposes in the out-of-sample prediction. The input RPs are used in the distance approximation step while the output RPs are used in the multilateration step. Therefore, the selection of input and output RPs must be considered as two different tasks.

Theoretically, in multilateration problems, more anchors[1] brings higher location accuracy [6, 7]. For that reason, in the problem of selecting output RPs, using all samples is acceptable. On the other hand, the problem of selecting input RPs can be considered as finding a possible good approximation to the true solution of the MMV linear system between distances computed from the input and output spaces, subject to sparsity and smoothness constraints. More specifically, selection of the input RPs must minimize the error and the row-diversity of the MMV linear system. That is, the points that represent non-zero rows of the coefficients matrix from the generated linear system are searched. Despite of this problem being generally regarded as NP-hard [8], a number of computational strategies have been developed to find solutions with low computational complexity [9, 10, 11].

A popular technique for finding a sparse solution of a MMV linear system is the extension of the focal underdetermined system solver (FOCUSS, [12]) for MMV linear systems problems with additive noise, called regularized MMV FOCUSS (regularized M-FOCUSS [9]). This algorithm employs an ℓ_p-norm-like diversity measure, where p is a user-defined parameter.

In order to adopt a sparse profile to the MLM, we propose the use of the regularized M-FOCUSS algorithm to select reference points in the input space. The proposed method, henceforth called regularized M-FOCUSS minimal learning machine (RMF-MLM), achieves results equivalent or superior to standard MLM techniques, besides generating less complex models. The results are confirmed by experiments with common databases from the literature.

The remainder of the paper is organized as follows: Section 2 briefly describes the regularized M-FOCUSS MLM. Section 3 reports the empirical assessment of the proposal and the conclusions are outlined in Section 4.

[1] In multilateration framework, a anchor (or node) is a sensor with a known position, used to approximate the position of some other nodes with unknown position.

2 Regularized M-FOCUSS minimal learning machine

Let $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n=1}^{N}$ be a data set, where $\mathcal{X} = \{\boldsymbol{x}_n\}_{n=1}^{N}$ and $\mathcal{Y} = \{\boldsymbol{y}_n\}_{n=1}^{N}$ are the input and the output data points, respectively, with $\boldsymbol{x}_n \in \mathbb{R}^D$ and $\boldsymbol{y}_n \in \mathbb{R}^S$. Furthermore, let $\mathcal{R} = \{\boldsymbol{r}_m\}_{m=1}^{M} \subseteq \mathcal{X}$ be the set of input RPs and $\mathcal{T} = \{\boldsymbol{t}_k\}_{k=1}^{K} \subseteq \mathcal{Y}$ be the set of output RPs. Moreover, let $\mathbf{D} \in \mathbb{R}^{N\times M}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{N\times K}$ be the distance matrices related to input and output, such that the m-th and k-th columns are respectively $[\|\boldsymbol{x}_1 - \boldsymbol{r}_m\|_2, \cdots, \|\boldsymbol{x}_N - \boldsymbol{r}_m\|_2]^T$ and $[\|\boldsymbol{y}_1 - \boldsymbol{t}_k\|_2, \cdots, \|\boldsymbol{y}_N - \boldsymbol{t}_k\|_2]^T$. The key idea behind MLM is the assumption of a linear mapping between $\mathbf{D}$ and $\boldsymbol{\Delta}$, giving rise to the following regression model:

$$\boldsymbol{\Delta} = \mathbf{D}\mathbf{B} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{M\times K}$ is the matrix of regression coefficients and $\boldsymbol{\varepsilon} \in \mathbb{R}^{N\times K}$ is a matrix of residuals. In the original proposal [2], an approximation $\hat{\mathbf{B}}$ of $\mathbf{B}$ can be achieved by the ordinary least squares estimate.

$$\hat{\mathbf{B}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{\Delta}. \tag{2}$$

Given a new input point $\boldsymbol{x}$, the approximation $\hat{\boldsymbol{\delta}} = [\hat{\delta}_1, \cdots, \hat{\delta}_K]$ of the distances between the output $\boldsymbol{y}$ of point $\boldsymbol{x}$ and the K output reference points is given by

$$\hat{\boldsymbol{\delta}} = [\|\boldsymbol{x} - \boldsymbol{r}_1\|_2, \cdots, \|\boldsymbol{x} - \boldsymbol{r}_M\|_2]\,\hat{\mathbf{B}}. \tag{3}$$

Therefore, an estimate $\hat{\boldsymbol{y}}$ of $\boldsymbol{y}$ can be obtained by the following problem:

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \sum_{k=1}^{K} \left((\boldsymbol{y} - \boldsymbol{t}_k)^T(\boldsymbol{y} - \boldsymbol{t}_k) - \hat{\delta}_k^2 \right)^2 \right\}, \tag{4}$$

which can be approached via any gradient-based optimization algorithm.

To select input RPs, we propose a new method called *regularized M-FOCUSS minimal learning machine* (RMF-MLM), which relies on a simultaneous sparse approximation method named regularized M-FOCUSS for the task of identifying which reference points are not relevant to the MLM's performance. This method employs an ℓ_p-norm-like diversity measure, where $p \in [0, 2]$ is a user-defined parameter that indicates the degree of sparsity. Initially, the RMF-MLM uses all the points as RPs (i.e. $\mathcal{R} = \mathcal{X}$ and $\mathcal{T} = \mathcal{Y}$) to compute the distance matrices $\mathbf{D}$ and $\boldsymbol{\Delta}$. After that, the M-FOCUSS is used to achieve an approximation of $\mathbf{B}$ by finding a local minimum of the following optimization problem

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}} \|\mathbf{D}\mathbf{B} - \boldsymbol{\Delta}\|_{\mathrm{F}}^2 + \lambda \sum_{m=1}^{M} \|\boldsymbol{b}_m\|_2^p \tag{5}$$

where $\lambda \geq 0$ is a trade-off parameter balancing estimation quality with diversity measure minimization, and $\boldsymbol{b}_m$ denotes the m-th row of $\mathbf{B}$.

The regularized M-FOCUSS MLM use the factored-gradient approach of [13, 14] to minimize (5). The algorithm iteractively updates $\mathbf{B}$ using the following steps:

$$\begin{aligned}
&\mathbf{W}^{[t+1]} = \operatorname{diag}\left\{(c_m^{[t]})^{1-p/2}\right\}, \text{ where } \; c_m^{[t]} = \left(\sum_{k=1}^{K}\left(b_{km}^{[t]}\right)^2\right)^{1/2}, \\
&\mathbf{Q}^{[t+1]} = \mathbf{D}^{[t+1]T}\left(\mathbf{D}^{[t+1]}\mathbf{D}^{[t+1]T} + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\Delta}, \text{ where } \; \mathbf{D}^{[t+1]} = \mathbf{D}\mathbf{W}^{[t+1]}, \qquad (6) \\
&\mathbf{B}^{[t+1]} = \mathbf{W}^{[t+1]}\mathbf{Q}^{[t+1]},
\end{aligned}$$

where t is the the current iterative step and $\mathbf{I} \in \mathbb{R}^{N\times N}$ is an identity matrix. The algorithm is terminated once a convergence criterion has been satisfied, e.g.,

$$\frac{\|\mathbf{B}^{[t+1]} - \mathbf{B}^{[t]}\|_{\mathrm{F}}}{\|\mathbf{B}^{[t]}\|_{\mathrm{F}}} < \tau, \qquad (7)$$

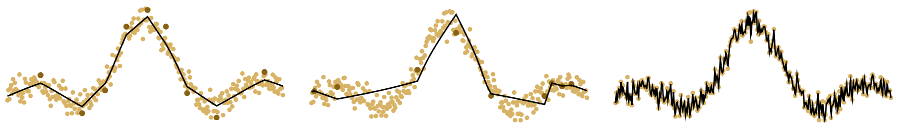
where τ is a the tolerance parameter. In our experiments, was chosen as 0.01.

In the regularized M-FOCUSS MLM, the choice of p is dictated by the speed of convergence and the sparsity of $\mathbf{B}$. Values of $p \leq 1$ encourages sparsity of solutions. If $p \to 0$, the regularized M-FOCUSS approaches a ℓ_0-norm-like.

3 Experiments

The performance of RMF-MLM is compared with two variants of the MLM, regarding the selection of input RPs. The first variant is the full MLM (FL-MLM), in which the set of input RPs is equal to the training set (i.e., $\mathcal{R} = \mathcal{X}$). The second variant is the random MLM (RN-MLM), where we randomly select M input RPs from the training data[2].

For a qualitative analysis, we have applied RMF-MLM, RN-MLM and FL-MLM to solve a toy problem that consists of 200 points in $\mathbb{R}^2$ regularly spaced in x, such that $x_n \in [-3\pi, 3\pi]$, $y_n = \frac{\sin x_n}{x_n} + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 0.1)$. In this experiment, we use for RN-MLM the number of input RPs found by the RMF-MLM using $p = 10^{-5}$ and $\lambda = 10^{-4}$. Fig. 1 shows the results of this experiment. Based on the Fig. 1, we can infer that RMF-MLM produced better output when



(a) $|\mathcal{R}| = 9$. (b) $|\mathcal{R}| = 9$. (c) $|\mathcal{R}| = 200$.

Fig. 1: Model outputs and number of input RPs for (b) RMF-MLM, (c) RN-MLM and (d) FL-MLM when applied to ART dataset.

[2] In all of this variants, the whole data outputs are used as RPs (i.e. $\mathcal{T} = \mathcal{Y}$)

compared to the other methods. In Fig. 1, one can see that both the FMR-MLM and RN-MLM use the same number of RPs. Moreover, the curve generated from the ON-MLM is smoother than the other models.

Tests with real-world benchmarking data sets were also carried out in this work. We used UCI data sets [15]: Motorcycle (MTR), Servo (SVR), CPU performance (CPU), Auto MPG (MPG), Boston Housing Data (BHD), Forest Fires (FFS), Concrete Compressive Strength (CCS), and Abalone (ABA), with dimensions 167×5, 209×7, 392×8, 506×14, 517×13, 1030×9, and 4177×9, respectively. In addition, a well-known artificial data set was also used in our simulations, the Motorcycle dataset, with dimension 94×2.

A nested 10-fold cross-validation was used in the experiments. The external 10-fold was used to estimate the performance metrics and the internal to adjust the user-selected parameters. The adjustment of the parameter M for the RN-MLM model was performed using the range of 5–95% (with a step size of 5%) of the available training samples and the distribution used to λ of the RMF-MLM are $\{10^{-5}, 10^{-4}, \cdots, 10^{-1}\}$. In Tab. 1, we report performance metrics for the aforementioned 10-fold cross-validation. We show the root mean squared error (RMSE) and the average of the number of the input RPs (#IRPs).

Table 1: Performance comparison – average values of root mean squared error (RMSE) and number of input reference points (#IRPs) for the tenfold cross-validation – with the RMF-MLM, RN-MLM and FL-MLM.

dataset	*metric*	RMF-MLM			RN-MLM	FL-MLM
		$p = 0.2$	$p = 0.4$	$p = 1$		
MTR	RMSE	23.35 ± 7.09	23.60 ± 6.67	23.62 ± 6.19	25.12 ± 5.90	30.77 ± 8.01
	#IRPs	9.70 ± 0.48	11.70 ± 0.82	26.60 ± 1.35	27.50 ± 8.03	84.60 ± 0.52
SRV	RMSE	0.56 ± 0.18	0.54 ± 0.20	0.54 ± 0.20	0.57 ± 0.25	0.54 ± 0.20
	#IRPs	38.00 ± 1.56	63.40 ± 2.22	139.20 ± 1.32	114.80 ± 21.80	150.30 ± 0.48
CPU	RMSE	43.37 ± 22.54	47.17 ± 31.59	45.08 ± 24.92	53.86 ± 39.11	45.18 ± 19.06
	#IRPs	12.10 ± 0.88	21.60 ± 6.65	67.30 ± 14.31	137.80 ± 48.30	188.10 ± 0.32
MGP	RMSE	2.59 ± 0.47	2.57 ± 0.46	2.57 ± 0.47	2.69 ± 0.46	2.63 ± 0.49
	#IRPs	49.40 ± 2.22	89.00 ± 2.87	238.20 ± 3.26	194.60 ± 70.50	352.80 ± 0.42
HSG	RMSE	2.95 ± 0.78	2.84 ± 0.79	2.80 ± 0.78	2.98 ± 0.82	2.77 ± 0.78
	#IRPs	85.00 ± 3.02	160.40 ± 5.50	380.90 ± 2.73	406.00 ± 57.80	455.40 ± 0.52
FFS	RMSE	60.11 ± 44.76	56.36 ± 46.50	57.47 ± 46.08	48.88 ± 47.58	55.72 ± 47.04
	#IRPs	35.20 ± 15.77	69.90 ± 29.12	230.70 ± 102.92	247.10 ± 117.67	465.30 ± 0.48
CCT	RMSE	6.06 ± 0.40	5.70 ± 0.47	5.52 ± 0.50	5.64 ± 0.50	5.03 ± 0.54
	#IRPs	173.20 ± 4.96	307.50 ± 3.78	722.90 ± 5.97	816.10 ± 76.21	927.00 ± 0.00
ABA	RMSE	2.13 ± 0.09	2.12 ± 0.10	2.12 ± 0.10	2.25 ± 0.35	2.22 ± 0.10
	#IRPs	220.20 ± 19.27	334.10 ± 17.90	752.50 ± 28.38	688.00 ± 20.01	3759.30 ± 0.48

As expected, the performance of the RMF-MLM was equivalent to or higher than the ones achieved by the RN-MLM and the FL-MLM for each evaluated data set. Particularly, the variants of the RMF-MLM achieved the best results (in terms of RMSE) in 4 of the data sets. With respect to the number of input RPs, our proposal achieved best results in all data sets. In other words, our proposal achieves errors that are comparable to others variants of MLM, but with a lower number of RPs. It is also important to notice that for most data sets the RMF-MLM achieved a low standard deviation.

ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, i6doc.com publ., ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/.

4 Conclusions

In this paper, we proposed an alternative algorithm to select input reference points for minimal learning machines for regression tasks based on a method for diversity measure minimization. The proposed approach was called regularized M-FOCUSS minimal learning machine (RMF-MLM). Three strategies of MLM input RPs selection were evaluated. On the basis of our experiments, we can state that RMF-MLM is a promising alternative to select input RPs, providing a competitive model while maintaining its simplicity.

References

[1] Amauri Holanda de Souza Júnior, Francesco Corona, Yoan Miche, Amaury Lendasse, Guilherme A. Barreto, and Olli Simula. Minimal learning machine: A new distance-based method for supervised learning. In *Advances in Computational Intelligence - 12th International Work-Conference on Artificial Neural Networks, IWANN 2013*, volume 7902 of *Lecture Notes in Computer Science*, pages 408–416. Springer, 2013.

[2] Amauri Holanda de Souza Junior, Francesco Corona, Guilherme De A. Barreto, Yoan Miché, and Amaury Lendasse. Minimal learning machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing*, 164:34–44, 2015.

[3] E. Niewiadomska-Szynkiewicz and M. Marks. Optimization schemes for wireless sensor network localization. *Applied Mathematics and Computer Science*, 19(2):291–302, 2009.

[4] Madson Luiz Dantas Dias, Lucas Silva de Sousa, Ajalmar R. da Rocha Neto, and Amauri H. Souza Júnior. Opposite neighborhood: a new method to select reference points of minimal learning machines. In *26th European Symposium on Artificial Neural Networks, ESANN 2018, Bruges, Belgium, April 25-27, 2018*, 2018.

[5] Átilla N. Maia, Madson Luiz Dantas Dias, João P. P. Gomes, and Ajalmar R. da Rocha Neto. Optimally selected minimal learning machine. In Hujun Yin, David Camacho, Paulo Novais, and Antonio J. Tallón-Ballesteros, editors, *Intelligent Data Engineering and Automated Learning - IDEAL 2018*, pages 670–678, 2018.

[6] Xin Li, Bei Hua, Yi Shang, and Yan Xiong. A robust localization algorithm in wireless sensor networks. *Frontiers Comput. Sci. China*, 2(4):438–450, 2008.

[7] Xinwei Wang, Ole Bischoff, Rainer Laur, and Steffen Paul. Localization in wireless ad-hoc sensor networks using multilateration with rssi for logistic applications. *Procedia Chemistry*, 1(1):461 – 464, 2009. Proceedings of the Eurosensors XXIII conference.

[8] D. P. Wipf and B. D. Rao. An empirical bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Processing*, 55(7-2):3704–3716, 2007.

[9] Shane F. Cotter, Bhaskar D. Rao, Kjersti Engan, and Kenneth Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Processing*, 53(7):2477–2488, 2005.

[10] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss. Algorithms for simultaneous sparse approximation. part I: greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[11] Jeffrey D. Blanchard, Michael Cermak, David Hanle, and Yirong Jing. Greedy algorithms for joint sparse recovery. *IEEE Trans. Signal Processing*, 62(7):1694–1704, 2014.

[12] Irina F. Gorodnitsky and Bhaskar D. Rao. Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Trans. Signal Processing*, 45(3):600–616, 1997.

[13] N. D. Rao and K. Kreutz-Delgado. An affine scaling methodology for best basis selection. *IEEE Trans. Signal Processing*, 47(1):187–200, 1999.

[14] Kenneth Kreutz-Delgado and Bhaskar D. Rao. Measures and algorithms for best basis selection. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, pages 1881–1884. IEEE, 1998.

[15] M. Lichman. UCI machine learning repository, 2013.