ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Opposite neighborhood: a new method to select reference points of minimal learning machines

Madson L. D. Dias[1], Lucas S. de Sousa[1],
Ajalmar R. da Rocha Neto[1], Amauri H. de Souza Júnior[1] *

1. Federal Institute of Ceará – Department of Teleinformatics
13 de Maio Avenue, 2081, Fortaleza, Ceará, Brazil

Abstract.
This paper introduces a new approach to select reference points in minimal learning machines (MLMs) for classification tasks. The MLM training procedure comprises the selection of a subset of the data, named reference points (RPs), that is used to build a linear regression model between distances taken in the input and output spaces. In this matter, we propose a strategy, named opposite neighborhood (ON), to tackle the problem of selecting RPs by locating RPs out of class-overlapping regions. Experiments were carried out using UCI data sets. As a result, the proposal is able to both produce sparser models and achieve competitive performance when compared to the regular MLM.

1 Introduction

The minimal learning machine (MLM, [1]) is a supervised learning algorithm that has recently been applied to a diverse range of problems, such as fault detection [2], ranking of documents [3], and robot navigation [4].

The basic operation of MLM consists in a linear mapping between the geometric configurations of points in the input space and the respective points in the output space. The geometric configuration is captured by two distance matrices (input and output), computed between the training/learning points and a subset of it whose elements are called reference points (RPs). The learning step in the MLM consists of fitting a linear regression model between these two distance matrices. In the test phase, given an input, the MLM predicts its output by first computing distances in the input space and then using the learned regression model to predict distances in the output space. Those distances are then used to provide an estimate to the output.

The determination of the RPs, including its quantity, is fundamental to the quality of the surface boundary generated by the MLM model. In this regard, the original formulation proposes a random sample from the training data as RPs. This paper tackles the problem of selecting reference points. We introduce a new approach, called opposite neighborhood MLM (ON-MLM), to select the RPs based on the Euclidean distance between samples of different classes. The rationale behind the proposal is to avoid selecting reference points in class-overlapping regions. We highlight that the ON method is inspired by a

*The authors would like to thank IFCE and CAPES for supporting their research.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

recent heuristic, called opposite maps (OM, [5]), proposed to building reduced-set for support vector machines (SVM, [6]) and least squares support vector machines (LSSVM, [7]) classifiers.

The remainder of the paper is organized as follows. Section 2 briefly describes the MLM. Section 3 introduces the ON-MLM. Section 4 reports the empirical assessment of the proposal and the conclusions are outlined in Section 5.

2 Minimal Learning Machine

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

$$\boldsymbol{\Delta}_y = \mathbf{D}_x \mathbf{B} + \mathbf{E} \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{M\times M}$ is the matrix of regression coefficients and $\mathbf{E} \in \mathbb{R}^{N\times K}$ is a matrix of residuals. Under the normal conditions where the number of selected reference points is smaller than the number of training points (i.e., $M < N$), the matrix $\mathbf{B}$ can be approximated by the usual least squares estimate

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

Given a new input point $\boldsymbol{x}$, the approximation $\hat{\boldsymbol{\delta}} = [\hat{\delta}_1, \cdots, \hat{\delta}_M]$ of the distances between the output $\boldsymbol{y}$ of point $\boldsymbol{x}$ and the M 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 minimization problem:

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \sum_{m=1}^{M} \left((\boldsymbol{y} - \boldsymbol{r}_m)^T (\boldsymbol{y} - \boldsymbol{r}_m) - \hat{\delta}_m^2 \right)^2 \right\}, \tag{4}$$

which can be approached via any gradient-based optimization algorithm. In the original paper, the regular MLM applies the Levenberg-Marquardt method [8].

For the classification case, where outputs $\boldsymbol{y}_n$ are represented using the 1-of-S encoding scheme [1]. It was showed in [9] that under the assumption that the classes are balanced, the optimal solution to Eq. (4) is given by

$$\hat{\boldsymbol{y}} = \boldsymbol{t}_{m^*}, \tag{5}$$

where $m^* = \arg\min_m \hat{\boldsymbol{\delta}}$. It means that output predictions for new incoming data can be carried out by simply selecting the output of the nearest reference point in the output space, estimated using the linear model $\hat{\mathbf{B}}$. This method was named Nearest Neighbor MLM (NN-MLM).

[1] A S-level qualitative variable is represented by a vector of S binary variables or bits, only one of which is *on* at a time. Thus, the j-th component of an output vector $\boldsymbol{y}$ is set to 1 if it belongs to class j and 0 otherwise.

3 Opposite Neighbohood Minimal Learning Machine

In this section, we describe our proposal named opposite neighborhood minimal learning machine (ON-MLM). As discussed, the MLM training procedure comprises *i*) the selection of reference points; and *ii*) the estimation of the coefficients of a multiresponse linear regression model. In this regard, the ON-MLM only tackles the problem of selecting reference points. In doing so, all the other steps remain the same as the original MLM, including the test procedure.

In a nutshell, our proposal relies on selecting RPs by locating RPs out of class-overlapping regions. The first step is to find and remove input data points from class-overlapping regions. The second one is to select the points in the separation region of the new reduced data set. To accomplish that, in the following we define the so called opposite neighborhood (ON) procedure.

Definition 1. *(Opposite neighborhood): Given a parameter* $K \in \mathbb{N}^+$ *and a data set* $\mathcal{D} = \{(\boldsymbol{x}_n, \boldsymbol{y}_n)\}_{n=1}^N$. *The result of the* K*-opposite-neighborhood is a subset* $\mathcal{Z} = ON(\mathcal{D}, K) \subseteq \mathcal{D}$, *given by*

$$\mathcal{Z} = \bigcup_{n=1}^{N} KNN(\boldsymbol{x}_n, \mathcal{M}_n, K), \tag{6}$$

where $\mathcal{M}_n = \{(\boldsymbol{x}_p, \boldsymbol{y}_p) \in \mathcal{D} \mid \boldsymbol{y}_p \neq \boldsymbol{y}_n\}$ *is the set of tuples containing input data points and their respective labels such that the labels differ from* $\boldsymbol{y}_n$; *and* $KNN(\cdot,\cdot,\cdot)$ *is the result of a* K*-nearest neighbor query [10].*

The application of the ON procedure to a database with class-overlapping regions may return a subset of points in those regions. On the other hand, when the ON procedure is applied to noiseless data, the resulting samples are those over the separation region among the classes.

As aforementioned, our proposal comprises two main steps. The first step is the removal of patterns located in class-overlapping regions. For this task, we use the ON procedure (with neighborhood size K) to detect the points to be removed. After that, the second steps aims to select the reference points over the reduced data set. This step is also made via the ON. However, in this case, the ON is applied only to the remaining samples (reduced set), with neighborhood size equal to 1. The Algorithm 1 depicts the pseudocode for the ON-MLM.

Algorithm 1 ON-MLM

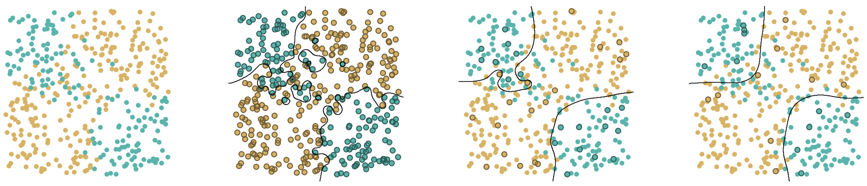
Require: training set $\mathcal{D}$ and the neighborhood size K
Ensure: reference points set $\mathcal{R}$
1: Removal of patterns in overlap region: $\mathcal{U} \leftarrow \mathcal{D} \setminus \mathrm{ON}(\mathcal{D}, K)$
2: Reference points selection: $\mathcal{R} \leftarrow \mathrm{ON}(\mathcal{U}, 1)$
3: **return** $\mathcal{R}$

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

4 Simulations and Discussion

The performance of ON-MLM is compared to two variants of the MLM, regarding the selection of RPs. The first variant is the full MLM (FL-MLM), in which the set of reference points is equal to the training set (i.e., $K = N$). The second variant is the random MLM (RN-MLM), where we randomly select K reference points from the training data. It corresponds to the original proposal. A combination of the k-fold cross-validation and holdout methods was used in the experiments. The holdout method with a 80% training and 20% test division was used to estimate the performance metrics. In Table 1 we report the performance metrics of each RP selection method.

For a qualitative analysis, we have also applied ON-MLM, RN-MLM and FL-MLM to solve an artificial problem. The toy problem, named simple checkerboard (SCB), consists of 1000 points in $\mathbb{R}^2$ taken from for black and white squares of a checkerboard (Figure 1 (a)).



(a) SCB data set (b) $|\mathcal{R}| = 235$ (c) $|\mathcal{R}| = 40$ (d) $|\mathcal{R}| = 19$

Fig. 1: Decision surface and number of RPs for (b) FL-MLM, (c) RN-MLM and (d) ON-MLM when applied to SCB dataset.

Based on the Fig. 1, we can infer that ON-MLM produced better decision boundaries when compared to the other methods. In the Figure 1, one can see that the number of RPs for ON-MLM is lower than the number of RPs for RN-MLM. Moreover, the decision boundary generated from the ON-MLM is more smoothed than the other models. Additionally, one can note in the qualitative analysis that the ON-MLM method avoids RPs on overlapping regions. Thus the decision boundaries are not overfitted.

Tests with real-world benchmarking data sets were carried out in this work. We used UCI data sets [11]; Hepatitis (HEP), Haberman's Survival (HAB), Vertebral Column Pathologies (VCP), Balance Scale (BLC), Breast Cancer Winconsin (BCW), Pima Indians Diabets (PID) and Human Immunodeficiency Virus protease cleavage (HIV) with 80, 306, 310, 625 ,688, 768, 6590 instances, respectively. In addition, three well-known artificial data sets were also used in our simulations, Two Moon (TMN), Ripley (RIP) and Banana (BNA), with 1001, 1249 and 5300 instances, respectively.

In our simulations, 80% of the data examples were randomly selected for training purposes and the remaining 20% of the examples were used for assessing the classifiers' generalization performances. We carried out 30 independent runs for each data set.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6. Available from http://www.i6doc.com/en/.

The adjustment of the parameter K for the RN-MLM and the ON-MLM model was performed using grid search combined with 10-fold cross-validation. The RPs were selected in the range of 5–100% (with a step size of 5%) of the available training samples. The classification error was used to choose the best value of K.

In Table 1, we report performance metrics for the aforementioned 30 independent runs. We also show the average of percentage of reduction in the # RPs compared to an FL-MLM model and the results of a statistical test.

Table 1: Performance comparison – Accuracy (ACC) and reduction percentage in comparison with the training set (RED) – with the FCM-MLM, RN-MLM and FL-MLM; and results of statistical tests. The symbols ✓ and ✗ with respect to the Friedman statistical test means fail to reject, and reject, respectively.

dataset	*metric*	OR-MLM	RN-MLM		FL-MLM	
HEP	ACC	**65.21** ± 10.34	64.17 ± 11.00	✓	61.67 ± 9.81	✓ ✓
	RED	**83.91** ± 4.77	47.81 ± 29.13			
HAB	ACC	**72.40** ± 4.89	71.97 ± 4.27	✓	68.09 ± 4.98	✗ ✗
	RED	**91.25** ± 1.94	80.20 ± 14.22			
VCP	ACC	**83.33** ± 5.38	82.58 ± 4.39	✓	82.15 ± 4.23	✓ ✓
	RED	**92.50** ± 2.00	56.51 ± 26.86			
BLC	ACC	99.84 ± 0.39	99.84 ± 0.44	✓	**100.00** ± 0.00	✓ ✓
	RED	**90.11** ± 1.04	81.64 ± 6.53			
BCW	ACC	**97.10** ± 1.45	96.98 ± 1.40	✓	96.96 ± 1.27	✓ ✓
	RED	**95.92** ± 0.55	62.61 ± 22.71			
PID	ACC	**74.78** ± 2.57	74.59 ± 2.58	✓	73.16 ± 2.38	✗ ✗
	RED	**88.90** ± 2.88	75.92 ± 16.10			
TMN	ACC	99.78 ± 0.28	99.82 ± 0.28	✓	**99.87** ± 0.22	✗ ✓
	RED	**97.62** ± 0.23	61.92 ± 20.72			
RIP	ACC	**90.00** ± 1.69	89.75 ± 1.77	✓	88.32 ± 1.61	✗ ✗
	RED	**97.08** ± 0.84	76.64 ± 18.83			
HIV	ACC	**86.72** ± 1.29	86.50 ± 1.30	✗	85.99 ± 1.14	✗ ✗
	RED	**99.52** ± 0.15	75.32 ± 23.16			
BNA	ACC	89.60 ± 0.74	**89.87** ± 0.81	✓	87.58 ± 0.89	✗ ✗
	RED	**96.42** ± 0.69	89.33 ± 2.54			

By analyzing the results and the the statistical hypothesis test carried out in Table 1 one can conclude that the performances of the ON-MLM were equivalent or even superior to those achieved by the RN-MLM and FL-MLM for each data sets evaluated. Moreover, one can also see that our proposal achieves sparse solutions.

5 Conclusions

In this paper, we propose an algorithm to select the reference points of the MLM for classification tasks based on the Euclidean distance between points of different classes. Three strategies of MLM reference point selection are evaluated. Our proposal called ON-MLM is able to obtain the RP subset for MLMs. The experimental results indicate that the ON-MLM works very well, providing a competitive classifier while maintaining its simplicity.

References

[1] 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.

[2] David N. Coelho, Guilherme De A. Barreto, Cláudio M. S. Medeiros, and José Daniel Alencar Santos. Performance comparison of classifiers in the detection of short circuit incipient fault in a three-phase induction motor. In *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions, CIES 2014, Orlando, FL, USA, December 9-12, 2014*, pages 42–48. IEEE, 2014.

[3] Alisson S. C. Alencar, Weslley L. Caldas, João Paulo Pordeus Gomes, Amauri H. de Souza, Paulo A. C. Aguilar, Cristiano Rodrigues, Wellington Franco, Miguel F. de Castro, and Rossana M. C. Andrade. Mlm-rank: A ranking algorithm based on the minimal learning machine. In *2015 Brazilian Conference on Intelligent Systems, BRACIS 2015, Natal, Brazil, November 4-7, 2015*, pages 305–309. IEEE, 2015.

[4] Leandro Balby Marinho, Jefferson S. Almeida, João Wellington M. Souza, Victor Hugo C. de Albuquerque, and Pedro Pedrosa Rebouças Filho. A novel mobile robot localization approach based on topological maps using classification with reject option in omnidirectional images. *Expert Syst. Appl.*, 72:1–17, 2017.

[5] Ajalmar R. da Rocha Neto and Guilherme A. Barreto. Opposite maps: Vector quantization algorithms for building reduced-set SVM and LSSVM classifiers. *Neural Processing Letters*, 37(1):3–19, 2013.

[6] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, New York, 1995.

[7] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

[8] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[9] Diego P. P. Mesquita, João P. P. Gomes, and Amauri H. Souza Junior. Ensemble of efficient minimal learning machines for classification and regression. *Neural Processing Letters*, pages 1–16, 2017.

[10] Nick Roussopoulos, Stephen Kelley, and Frédéic Vincent. Nearest neighbor queries. In Michael J. Carey and Donovan A. Schneider, editors, *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, San Jose, California, May 22-25, 1995.*, pages 71–79. ACM Press, 1995.

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