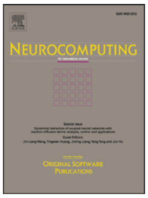
Neurocomputing 248 (2017) 11–18

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Euclidean distance estimation in incomplete datasets

Diego P.P. Mesquita[a], João P.P. Gomes[a,*], Amauri H. Souza Junior[c], Juvêncio S. Nobre[b]

[a] Federal University of Ceará, Department of Computer Science, Fortaleza, CE, Brazil
[b] Federal University of Ceará, Department of Statistics and Applied Mathematics, Fortaleza, CE, Brazil
[c] Federal Institute of Ceará, Department of Computer Science, Maracanaú, CE, Brazil

ARTICLE INFO

Article history:
Received 1 July 2016
Revised 20 October 2016
Accepted 17 December 2016
Available online 8 March 2017

Keywords:
Missing data
Euclidean distance
Minimal Learning Machine

ABSTRACT

This paper proposes a method to estimate the expected value of the Euclidean distance between two possibly incomplete feature vectors. Under the Missing at Random assumption, we show that the Euclidean distance can be modeled by a Nakagami distribution, for which the parameters we express as a function of the moments of the unknown data distribution. In our formulation the data distribution is modeled using a mixture of Gaussians. The proposed method, named Expected Euclidean Distance (EED), is validated through a series of experiments using synthetic and real-world data. Additionally, we show the application of EED to the Minimal Learning Machine (MLM), a distance-based supervised learning method. Experimental results show that EED outperforms existing methods that estimate Euclidean distances in an indirect manner. We also observe that the application of EED to the MLM provides promising results.

1. Introduction

Data completeness is a major assumption of most machine learning methods. In real world problems, however, several data instances may suffer from unobserved/missing attributes. This issue, referred to as missing/incomplete data problem, may happen due to a variety of reasons such as sensor problems, device malfunction and operator mistakes [1]. The simplest way to deal with missing data consists of removing the instances with missing attributes (listwise deletion) from the dataset. Even though this approach may work in some cases, discarding data samples usually leads to loss of important information to build a learning model [2]. Another widely used approach is to perform a pre-processing step of missing data imputation. After filling the missing entries, any conventional learning method can be used. Examples of such an approach can be found in [3–5] and [6].

According to Acuña and Rodrigues in [7], problems with more than 5% of missing samples may require sophisticated handling methods. In such situations, good results can be achieved by not considering the imputation as a separate step. Rather, it is possible to design a learning method that can handle incomplete data in its formulation. By doing so, the inherent uncertainty of the imputation process is taken into account and it has shown to be beneficial in many cases [8].

Computing the Euclidean distance is a key part in many machine learning methods, such as *k*-nearest neighbors [9], *k*-means [10] and Learning Vector Quantization [11]. In this paper, we propose a strategy to estimate pairwise Euclidean distance between vectors with missing values. The method, named Expected Euclidean Distance (EED) estimation, aims to directly determine the expected value of the Euclidean distance between two potentially incomplete vectors under the assumption that the distances are Nakagami-distributed [12]. The expected value of the Euclidean distance has a closed form solution that depends only on the two parameters of the Nakagami distribution. We show that these parameters can be expressed in terms of the non-central moments of the unknown data distribution. To model the data distribution, we adopt a Gaussian mixture distribution whose parameters are estimated using the maximum likelihood method [13].

To assess the effectiveness of the Expected Euclidean Distance estimation method in the context of supervised learning, we describe the application of the EED in the recently proposed Minimal Learning Machine (MLM, [14]). MLM is a distance-based supervised learning algorithm based on the idea of the existence of a mapping between the geometric configurations of points in the input and output space. These geometric configurations are expressed in terms of distances matrices between data samples. The proposed variant of the MLM for missing data uses the EED to build the input distance matrix. Both EED and its application to the MLM are benchmarked using artificial and real-world datasets. Based on the

[*] Corresponding author.
E-mail addresses: diegoparente@lia.ufc.br (D.P.P. Mesquita), jpaulo@lia.ufc.br, jpaulopg@gmail.com (J.P.P. Gomes), amauriholanda@ifce.edu.br (A.H. Souza Junior), juvencio@ufc.br (J.S. Nobre).

http://dx.doi.org/10.1016/j.neucom.2016.12.081

experiments, we show that EED (i) represents a promising alternative to compute distances in cases where missing data is an issue; and (ii) it can be successfully applied to distance-based learning methods.

The remainder of this paper is structured as follows. Section 2 explores related works on missing data and distance estimation. Section 3 states the problem of estimating Euclidean distances on datasets with missing values and introduces the proposed solution. In Section 4, we propose a MLM variant for missing data. Experiments on synthetic and real-world datasets are presented in Section 5. Conclusions and future work are given in Section 6.

2. Related work

The problem of missing data can be classified according to the missingness mechanism. Little and Rubin in [15] characterize the mechanisms as Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR). In MCAR, the missingness of a value is independent of its value and any other component of the dataset. A less restrictive assumption is the MAR. MAR is a more likely to happen in reality [16] and states that the missingness of a component is independent of the value itself but its value can be estimated using the observed components. Finally, the MNAR mechanism characterizes the situation in which the probability of missingness is related to the value of the component itself.

Under the assumption that the components are MAR, several works have been proposed to estimate distances in datasets with missing values. A popular method is the Partial Distance Strategy (PDS, [17]). In the PDS, a distance is computed using only the known components of each vectors. After that, the distance is proportionally scaled to account for the missing values. This strategy is known to underestimate distances [2] and it is impossible to apply if the vectors do not have common known components.

Eirola et al. in [2] proposed the Expected Square Distance (ESD) method to calculate the expected value of the squared distance between vectors with missing values. Although results were promising, the ESD assumed that the data was normally distributed. An improved version of this method is presented in [1], where the dataset is modeled with a Gaussian Mixture Model. Both works in [2] and [1] are not limited the Euclidean distance.

It is important to notice that the problem of Euclidean distance estimation is not addressed in [2] or [1]. Instead, the Euclidean distance can be estimated by computing the squared root of the expected squared Euclidean distance. This procedure may affect the performance of distance-based learning methods, since taking expectation over the square root of a random variable is not the same as computing the square root of the expected value. Although using the expected squared Euclidean distance instead of the expected Euclidean distance may be a valid option in some situations, this is not truth for methods that rely on the use of a proper distance metric. It is straightforward to verify that the squared distance does not satisfy the triangle inequality and thus does not correspond to a metric.

3. The expected euclidean distance

Consider two D-dimensional vectors $X_i = (x_{i,1}, \dots, x_{i,D})^T$ and $X_j = (x_{j,1}, \dots, x_{j,D})^T$, the Euclidean distance η between X_i and X_j is given by

$$\eta = z^{1/2} = \sqrt{\sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2}, \tag{1}$$

where $z = \|X_i - X_j\|_2^2 = \sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2$ is the squared distance between X_i and X_j.

We are interested in estimating η when any of the vectors $X_i, X_j \in \mathcal{X}$ have one or more missing component. Specifically, we consider the case that unobserved values in X_i and X_j are MAR [15]. Furthermore, we assume X_i and X_j are independent.

3.1. Formulation

Note that η can be considered a random variable since it is a transform of X_i and X_j, for which the missing entries can be modeled as such. Given η to be non-negative with Probability Density Function (PDF) $p(\eta)$, its expected value is given by

$$\mathrm{E}[\eta] = \int_0^{+\infty} p(\eta)\eta \, d\eta. \tag{2}$$

To solve the integral above, a statistical model for $p(\eta)$ is needed. From the definition of z, we have

$$z = \sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2 = \sum_{d=1}^{D} \phi_d^2, \tag{3}$$

Considering that any of the components $x_{i,\ d}$ or $x_{j,\ d}$ is missing, z can be characterized as a sum of squared random variables. According to [18], the distribution of a squared random variable ϕ^2 is said to be Gamma if the pdf of ϕ is given by:

$$p(\phi) = h(\phi)|\phi|^{2\alpha-1} exp\{-\beta\phi^2\}, \tag{4}$$

where α and β are the parameters of the distribution, and $\forall\phi :$ $h(\phi) + h(-\phi) = \zeta$ with ζ a constant.

It is worth noting that distributions with various characteristics can be written as specified in Eq. (4). Examples are the Gaussian distribution, the skew normal [19], Kotz-type distributions that are bi-modal and may have light tails, as well as other heavy- and light-tail distributions obtained similarly to the skew normal distribution (see [18,20] and [21]). This diversity of distributions makes the modeling of z as a sum of Gamma distributed random variables, a reasonable assumption. Finally, in [22] the authors showed that the sum of Gamma random variables can be approximated with a Gamma distribution.

Considering that z can be successfully modeled using a Gamma distribution, it is reasonable to choose the Nakagami [12] distribution for η. According to its definition, a random variable $\phi \sim$ Nakagami$(m,\ \Omega)$ can be obtained by taking the square root of $\varphi \sim$ Gamma$(\alpha,\ \beta)$. The Nakagami distribution is a function of two parameters (shape and spread) and is often used in communications theory to model scattered signals reaching a receiver by multiple paths [23]. Under the assumption that $\eta \sim$ Nakagami$(m,\ \Omega)$, the expected value of η is given by:

$$\mathrm{E}[\eta] = \frac{\Gamma(m + \frac{1}{2})}{\Gamma(m)} \left(\frac{\Omega}{m}\right)^{\frac{1}{2}}, \tag{5}$$

where m and Ω are respectively the shape and spread parameters of the Nakagami distribution and Γ is the Gamma function. The parameters m and Ω can be written as functions of the mean and variance of z according to Eq. (6).

$$m = \frac{\mathrm{E}^2[z]}{\mathrm{Var}[z]}, \ \Omega = E[z]. \tag{6}$$

Under the assumption of independence between missing entries, the problem of estimating E[z] can be approached using the results from [2] and [1], in which expected squared distances are expressed in the form

$$\mathrm{E}[z] = \sum_{d \notin M_i \cup M_j} (x_{i,d} - x_{j,d})^2 + \sum_{d \in M_j \setminus M_i} \mathrm{E}[(x_{i,d} - x_{j,d})^2] + \sum_{d \in M_i \setminus M_j} \mathrm{E}[(x_{i,d} - x_{j,d})^2] + \sum_{d \in M_i \cap M_j} \mathrm{E}[(x_{i,d} - x_{j,d})^2], \quad (7)$$

where $M_i, M_j \subseteq \{1, \ldots, D\}$ denote the sets of indexes of the missing components of X_i and X_j, respectively. Since all of the terms in Eq. (7) can be expanded to yield

$$\begin{aligned}\mathrm{E}[(x_{i,d} - x_{j,d})^2] &= \mathrm{E}[x_{i,d}^2] + \mathrm{E}[x_{j,d}]^2 - 2\mathrm{E}[x_{i,d}]\mathrm{E}[x_{j,d}] \\ &= \mathrm{E}[x_{i,d}^2] - \mathrm{E}[x_{i,d}]^2 + \mathrm{E}[x_{j,d}^2] - \mathrm{E}[x_{j,d}]^2 + \mathrm{E}[x_{i,d}]^2 \\ &\quad + \mathrm{E}[x_{j,d}]^2 - 2\mathrm{E}[x_{i,d}]\mathrm{E}[x_{j,d}] \\ &= (\mathrm{E}[x_{i,d}] - \mathrm{E}[x_{j,d}])^2 + \mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}]\end{aligned}$$

the expected squared distance $\mathrm{E}[z]$ can be compactly written as

$$\mathrm{E}[z] = \sum_{d=1}^{D} (\mathrm{E}[x_{i,d}] - \mathrm{E}[x_{j,d}])^2 + \mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}]. \quad (8)$$

Using a similar procedure, we can compute the variance of z. By expanding the expression $\mathrm{Var}[z] = \mathrm{E}[z^2] - \mathrm{E}[z]^2$, we have:

$$\begin{aligned}\mathrm{Var}[z] &= \mathrm{Var}\left[\sum_{d=1}^{D} (x_{i,d} - x_{j,d})^2\right] \\ &= \sum_{d=1}^{D} \mathrm{Var}\big[(x_{i,d} - x_{j,d})^2\big] \\ &\quad + \sum_{d=1}^{D} \sum_{l=d+1}^{D} \mathrm{Cov}\big[(x_{i,d} - x_{j,d})^2, (x_{i,l} - x_{j,l})^2\big] \quad (9)\end{aligned}$$

under the same independence assumptions taken in [1,2], the covariance terms go to zero and we are left with:

$$\begin{aligned}\mathrm{Var}[z] &= \sum_{d=1}^{D} \mathrm{Var}\big[(x_{i,d} - x_{j,d})^2\big] \\ &= \sum_{d=1}^{D} \mathrm{E}\big[(x_{i,d} - x_{j,d})^4\big] - \mathrm{E}\big[(x_{i,d} - x_{j,d})^2\big]^2 \\ &= \left(\sum_{d=1}^{D} \mathrm{E}\big[x_{i,d}^4 + x_{j,d}^4 - 4x_{i,d}^3 x_{j,d} - 4x_{i,d} x_{j,d}^3 + 6x_{i,d}^2 x_{j,d}^2\big] \right) \\ &\quad - \sum_{d=1}^{D} \mathrm{E}\big[(x_{i,d} - x_{j,d})^2\big]^2 \quad (10)\end{aligned}$$

It is possible to notice that both Eqs. (8) and (10) can be expressed in terms of non-central moments of X_i and X_j. Such moments can be estimated by imposing a distribution from which X_i and X_j are drawn and estimating the parameters of such distribution.

3.2. Modeling the data with a Gaussian mixture distribution

In the present work, we assume that X can be modeled by a weighted mixture of Gaussian distributions. The Gaussian Mixture Model (GMM) for datasets with missing values was proposed in [24] and used in many works such as [1].

For a vector $X_n \in \mathbb{R}^D$ the GMM PDF can be defined by the following equation:

$$p(X_n) = \sum_{c=1}^{C} w^{(c)} \mathcal{N}(X_n | \mu^{(c)}, \Sigma^{(c)}) \quad (11)$$

where $\mathcal{N}(X_n | \mu^{(c)}, \Sigma^{(c)})$ denotes the PDF of the multivariate normal distribution and $\{w^{(c)}\}_{c=1}^{C}$ is a set of weights such that $w^{(c)} \in [0, 1]$ and $\sum_{c=1}^{C} w^{(c)} = 1$. The weights $w^{(c)}$ combine C Normal distributions and assure that $p(X_n)$ is a PDF. In [24], the parameters $\mu^{(c)}$, $\Sigma^{(c)}$ and $w^{(c)}$ are found trough an Expectation-Maximization procedure.

For an arbitrary vector X_n with missing components denoted by $X_{n,\,M}$ and observed components denoted by $X_{n,\,O}$, where M and O are the sets of indexes of observed and missing component values respectively, the conditional mean $\tilde{\mu}_n^{(c)}$ and conditional covariance matrix $\tilde{\Sigma}_n^{(c)}$ of the cth Gaussian in the GMM are given by :

$$\tilde{\mu}_n^{(c)} = \mu_M^{(c)} + \Sigma_{MO}^{(c)} (\Sigma_{OO}^{(c)})^{-1} (X_{n,O} - \mu_O^{(c)}) \quad (12)$$

$$\tilde{\Sigma}_n^{(c)} = \Sigma_{MM}^{(c)} - \Sigma_{MO}^{(c)} (\Sigma_{OO}^{(c)})^{-1} \Sigma_{OM}^{(c)} \quad (13)$$

and the moments noncentral of the GMM distribution are given by:

$$\mathrm{E}[x_{n,d}] = \sum_{c=1}^{C} w^{(c)} \tilde{\mu}_{n,d}^{(c)} \quad (14)$$

$$\mathrm{E}[x_{n,d}^2] = \sum_{c=1}^{C} w^{(c)} \left([\tilde{\mu}_{n,d}^{(c)}]^2 + \tilde{\Sigma}_{n,d}^{(c)} \right) \quad (15)$$

$$\mathrm{E}[x_{n,d}^3] = \sum_{c=1}^{C} w^{(c)} \left([\tilde{\mu}_{n,d}^{(c)}]^3 + 3\tilde{\mu}_{n,d}^{(c)} \tilde{\Sigma}_{n,d}^{(c)} \right) \quad (16)$$

$$\mathrm{E}[x_{n,d}^4] = \sum_{c=1}^{C} w^{(c)} \left([\tilde{\mu}_{n,d}^{(c)}]^4 + 6[\tilde{\mu}_{n,d}^{(c)}]^2 \tilde{\Sigma}_{n,d}^{(c)} + 3[\tilde{\Sigma}_{n,d}^{(c)}]^2 \right), \quad (17)$$

where $\tilde{\mu}_{n,d}^{(c)}$ is the dth element of vector $\tilde{\mu}_n^{(c)}$ and $\tilde{\Sigma}_{n,d}^{(p)}$ is the dth element of the main diagonal of matrix $\tilde{\Sigma}_n^{(p)}$. Recall the $\mathrm{Var}[z]$, Eq. (10), depends only on the noncentral moments of X_i and X_j. Therefore, for this specific setup, it can be further simplified to:

$$\begin{aligned}\mathrm{Var}[z] &= \sum_{d=1}^{D} 4(\mathrm{E}[x_{i,d}] - \mathrm{E}[x_{j,d}])^2 (\mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}]) \\ &\quad + 2(\mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}])^2. \quad (18)\end{aligned}$$

Therefore, the EED method can be described as in Algorithm 1, in which the GMM that is taken as input comprises N components and the weight, mean vector and covariance matrix of the cth component are represented, respectively, by $w^{(c)}$, $\mu^{(c)}$ and $\Sigma^{(c)}$. The parameters of the GMM can be estimated via maximum likelihood with the aid of an Expectation Maximization algorithm take includes the unobserved entries in the set of latent variables.

4. Application to the Minimal Learning Machine

Minimal Learning Machine is a recently proposed distance-based supervised learning algorithm. The basic idea behind the MLM is the existence of a linear mapping between distances taken from the input and output spaces [14]. The MLM can be divided into two steps: distance regression (training) and output estimation (test).

Consider a training set $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}\}$ with N examples, such that $\mathcal{X} = \{X_i\}_{i=1}^{N}$ and $\mathcal{Y} = \{Y_i\}_{i=1}^{N}$ are respectively a set of D-dimensional input points and their S-dimensional outputs. The distance regression step starts by randomly selecting a subset of the training data, named reference points. The set of inputs of the chosen reference points is denoted by $\mathcal{R} = \{M_k\}_{k=1}^{K}$ and its corresponding outputs $\mathcal{T} = \{T_k\}_{k=1}^{K}$, where K is the number of reference points. Based on that, we can define $D_x \in \mathbb{R}^{N \times K}$ such that its kth column $D_x(\mathcal{X}, M_k)$ contains the distances $\{d(X_i, M_k)\}_{i=1}^{N}$ between the N input points

Algorithm 1 EED.

```
Require: $X_i, X_j \in \mathcal{X}$ and $GMM$ previously estimated from $\mathcal{X}$ comprising $C$ components.
Ensure: Estimate $\hat{\eta}$ of the expected euclidean distance between $X_i$ and $X_j$
    {Initialization.}
$M_i \leftarrow \{d : X_{i,d}$ is missing $\}$
$O_i \leftarrow \{1, \cdots, D\} \setminus M_i$
$M_j \leftarrow \{d : X_{j,d}$ is missing $\}$
$O_j \leftarrow \{1, \cdots, D\} \setminus M_j$
$\hat{\eta} \leftarrow 0$
    {Condition each of the GMM components on the observed values of both $X_i$ and $X_j$.}
for $c = 1 \cdots C$ do
    $\tilde{\mu}_i^{(c)} \leftarrow \mu_{M_i}^{(c)} + \Sigma_{M_i O_i}^{(c)} (\Sigma_{O_i O_i}^{(c)})^{-1} (X_{i,O_i} - \mu_{O_i}^{(c)})$
    $\tilde{\Sigma}_i^{(c)} \leftarrow \Sigma_{M_i M_i}^{(c)} - \Sigma_{M_i O_i}^{(c)} (\Sigma_{O_i O_i}^{(c)})^{-1} \Sigma_{O_i M_i}^{(c)}$
    $\tilde{\mu}_j^{(c)} \leftarrow \mu_{M_j}^{(c)} + \Sigma_{M_j O_j}^{(c)} (\Sigma_{O_j O_j}^{(c)})^{-1} (X_{j,O_j} - \mu_{O_j}^{(c)})$
    $\tilde{\Sigma}_j^{(c)} \leftarrow \Sigma_{M_j M_j}^{(c)} - \Sigma_{M_j O_j}^{(c)} (\Sigma_{O_j O_j}^{(c)})^{-1} \Sigma_{O_j M_j}^{(c)}$
end for
    {Compute padded conditional mean vectors and conditional covariance matrices of $X_i - X_j$ for each GMM component.}
for $c = 1 \cdots C$ do
    $\bar{\mu}^{(c)} \leftarrow [\mathbf{0}]^{\mathbf{d}}$
    $\bar{\mu}_{O_i}^{(c)} \leftarrow \bar{\mu}_{O_i}^{(c)} + X_{i,O_i}$
    $\bar{\mu}_{O_j}^{(c)} \leftarrow \bar{\mu}_{O_j}^{(c)} + X_{j,O_j}$
    $\bar{\mu}_{M_i}^{(c)} \leftarrow \bar{\mu}_{M_i}^{(c)} + \tilde{\mu}_i^{(c)}$
    $\bar{\mu}_{M_j}^{(c)} \leftarrow \bar{\mu}_{M_j}^{(c)} + \tilde{\mu}_j^{(c)}$
    $\hat{\Sigma}^{(c)} \leftarrow [\mathbf{0}]^{\mathbf{d \times d}}$
    $\hat{\Sigma}_{M_i,M_i}^{(c)} \leftarrow \hat{\Sigma}_{M_i,M_i}^{(c)} + \hat{\Sigma}_i^{(c)}$
    $\hat{\Sigma}_{M_j,M_j}^{(c)} \leftarrow \hat{\Sigma}_{M_j,M_j}^{(c)} + \hat{\Sigma}_j^{(c)}$
end for
    {Compute $\hat{\eta}$.}
for $d = 1 \cdots D$ do
    $m \leftarrow \sum_{c=1}^{C} w^{(c)} \bar{\mu}_d^{(c)}$
    $s \leftarrow \sum_{c=1}^{C} w^{(c)} ([\bar{\mu}_d^{(c)}]^2 + \bar{\Sigma}_{d,d}^{(c)})$
    $v \leftarrow s - m^2$
    $\hat{\eta} \leftarrow \hat{\eta} + 4m^2 v + 2v^2$
end for
```

X_i and the kth reference point M_k. Analogously, define $\Delta_y \in \mathbb{R}^{N \times K}$ in such a way that its kth column $\Delta_y(\mathcal{Y}, T_k)$ contains the distances $\{\delta(Y_i, T_k)\}_{i=1}^{N}$ between the N output points Y_i and the output T_k of the kth reference point.

Assuming that there exists a linear map g between the row space of D_x and the row space of Δ_y, it gives rise to the linear regression model given by

$$\Delta_y = D_x B + E, \tag{19}$$

where $E \in \mathbb{R}^{N \times K}$ represents residuals, and B denotes the linear map or parameters of the multiresponse regression model. It turns out that B can be estimated using ordinary least squares:

$$\hat{B} = (D_x^T D_x)^{-1} D_x^T \Delta_y. \tag{20}$$

For an input test point X whose distances from the K reference input points $\{M_k\}_{k=1}^{K}$ are in the vector $D(X, \mathcal{R}) = [d(X, M_1) \ldots d(X, M_K)]$, the corresponding estimated distances between its unknown output Y and the known outputs $\{T_k\}_{k=1}^{K}$ of the reference points are

$$\hat{\Delta}(Y, \mathcal{T}) = D(X, \mathcal{R}) \hat{B}. \tag{21}$$

The output estimation step consist of estimating the output Y from the outputs of all the reference points $\{T_k\}_{k=1}^{K}$ and the distance estimates $\hat{\Delta}(Y, \mathcal{T})$. The location of Y is the one that minimizes the objective function in Eq. (22), and it can be achieved using any gradient-based optimization algorithm.

$$J(Y) = \sum_{k=1}^{K} \left((Y - T_k)^T (Y - T_k) - \hat{\delta}^2(Y, T_k) \right)^2. \tag{22}$$

Similarly to other machine learning methods, MLM rely on the assumption that feature vectors have a fixed dimension and none of its features are missing. In this work, we propose a variant of the MLM for datasets with missing values, the Expected Euclidean Distance Minimal Learning Machine (EED-MLM). In this method, we implement the EED strategy to estimate the expected values of the input distance matrix D_x. After that, the estimated D_x is used to calculate $\hat{B}$ according to Eq. (20). The output estimation step of EED-MLM remains as originally proposed in the MLM.

5. Experiments and results

To assess the performance of the proposed method, we conducted four experiments with both synthetic and real world data. The first three experiments aim to evaluate how well the EED method reconstruct Euclidean distances. In the latter, we evaluate the usage of the EED in the design of the MLM for incomplete data.

The EED method is compared to the Conditional Mean Imputation (CMI, [24]) and the Expected Squared Distance (ESD,[1]). In CMI the missing entries are filled with its expected value conditioned to the observed values of the same instance, then the Euclidean distance is calculated using the imputed values. The ESD method estimates the expected squared distance and the Euclidean distance is computed by taking its squared root. A relation between CMI, ESD and EED can be seen in Eq. (23).

$$\begin{aligned}
\eta^{EED}(X_i, X_j) &= \mathrm{E}\left[\sqrt{\|X_i - X_j\|_2^2} \,\middle|\, X_{i,O}, X_{j,O} \right] \\
\eta^{ESD}(X_i, X_j) &= \sqrt{\mathrm{E}\left[\|X_i - X_j\|_2^2 \,\middle|\, X_{i,O}, X_{j,O} \right]} \\
\eta^{CMI}(X_i, X_j) &= \sqrt{\|\mathrm{E}[X_i - X_j | X_{i,O}, X_{j,O}]\|_2^2}
\end{aligned} \tag{23}$$

5.1. Univariate normal data with known parameters

In this experiment, we are interested in verifying the impact of uncertainty on the estimation of Euclidean distances. We assume that the missing samples come from a normal distribution, i.e., $X_i \sim \mathcal{N}(2, \sigma^2)$, where uncertainty is represented by σ^2. We want then estimate Euclidean distances η_i between X_i and a fixed sample $X_j = 3$. In this experiment we do not use GMM to estimate the distribution of the data, rather we consider σ^2 as known. Consequently, the performances of the methods do not depend on the accuracy in estimating the missing entry of the samples X_i. To obtain a benchmark, we compute a Monte Carlo (MC) estimate of η_i by averaging the Euclidean distances computed for 10^8 draws of X_i. It is important to highlight that, as we used MC results as a baseline, the performance of the other methods are as good as they are similar to MC. The impact of the uncertainty is verified by repeating this procedure while increasing the uncertainty in X_i which is encoded in σ^2. Table 1 shows the averaged Euclidean distance computed by each method.

As can be noticed, CMI results are independent of σ^2. This is expected since CMI does not account for the uncertainty of the estimates in its formulation. In ESD, as shown in [2], the variance of the estimates is taken into account and it provides more accurate

Table 1
Euclidean distance estimates.

σ^2	MC	CMI	ESD	EED
10^{-2}	1	1	1.005	1
10^{-1}	1.0002	1	1.0488	1.0045
10^{0}	1.1667	1	1.4142	1.1866
10^{1}	2.6481	1	3.3166	2.6505
10^{2}	8.019	1	10.0499	8.0188

Table 2
RMSE for the synthetic dataset. The symbols ✓ and ✗ indicate the result of the hypothesis test (✓ fail to reject, and ✗ reject).

Method	**RMSE**				
	Missing Data (%)				
	10	20	30	40	50
CMI	0.9876 ✗	0.9902 ✗	0.9927 ✗	0.9952 ✗	0.9989 ✗
ESD	0.9872 ✓	0.9895 ✓	0.9916 ✗	0.9936 ✗	0.9964 ✗
EED	0.9872	0.9894	0.9915	0.9934	0.9962

Table 3
Relative success rate of EED vs. CMI and ESD. The symbols ✓ and ✗ indicate the result of the hypothesis test (✓ fail to reject, and ✗ reject).

Method	**Relative Success Rate**				
	Missing Data (%)				
	10	20	30	40	50
CMI	0.90 ✗	0.93 ✗	1.00 ✗	1.00 ✗	1.00 ✗
ESD	0.80 ✗	0.83 ✗	0.86 ✗	0.76 ✗	0.86 ✗

Table 4
Description of the datasets.

Dataset	Size	Features
Automobile Price (AP)	159	15
Auto MPG (AM)	392	7
Boston Housing (BH)	506	13
Concrete Compression (CC)	1030	8
Servo (SV)	167	4

results than those obtained with CMI. Even though the ESD performance is significantly better than the CMI one, it degrades as σ^2 increase. In contrast to that, EED maintains a steady performance and reports the best results throughout the experiment.

5.2. Multivariate normal data with known parameters

This experiment aims to assess the performance of EED as the amount of missing data increases. Let X_i and X_j be drawn from a Multivariate Normal with parameters

$$\mu = \begin{bmatrix} -0.3 \\ 0.1 \\ 2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 0.40 & 0.15 & 0.25 \\ 0.15 & 0.25 & 0.10 \\ 0.25 & 0.10 & 0.30 \end{bmatrix}. \tag{24}$$

We created a dataset with 100 samples from the specified distribution and deleted some components of each sample. The components are deleted in a way to assure data is MAR. Once again we did not use the GMM to estimate the distribution of the data. All methods used the parameters specified in Eq. (24) to estimate the Euclidean distances.

The percentage of instances with missing components varied from 10% to 50%, and at each level we computed Euclidean distances between all pairs of vectors using EED, CMI and ESD. Table 2 shows the performances in terms of Root Mean Squared Error (RMSE) averaged over 30 independent runs. In order to ensure the statistical significance of the results achieved by the EED against the CMI and ESD, we run the Wilcoxon signed-rank test with a 5% significance level. In this hypothesis test, the null hypothesis state that the difference between quantities comes from a distribution with zero location parameter.

As can be noticed CMI had the worst overall performance, emphasizing the importance of accounting for uncertainty when estimating Euclidean distances in missing data. Additionally, it is also possible to notice that EED outperforms ESD as the amount of missing components increases.

Another interesting aspect that has to be observed when analyzing the performance of EED is the consistency of its results. In this work, consistency is measured by the so-called relative success rate – the rate in which a method outperforms other one. For example, if EED provides smaller RMSEs than the ESD method in 80 out of 100 repetitions, it means that EED is 0.8 consistently better than ESD, providing relative success rate of 0.8. The entries on Table 3 show the relative success rate for the proposed EED in comparison to CMI and ESD. The statistical significance of these results was tested using the chi-squared test with a significance level of 5%.

Once again EED had the best overall performance, achieving more accurate estimates in most of the executions. Comparing CMI to EED we can notice that the performance gap increases with the number of missing data. On the other hand, the difference between ESD and EED is roughly constant, although EED is superior in, at least, 76% of the cases.

5.3. Distance estimation on real-world data

We evaluate the performance of EED on real-world datasets. For that purpose, six datasets were selected from the UCI Machine Learning Repository [25]. Details about the number of features and dimensionality of each dataset are available in Table 4.

In this experiment, we compare EED to CMI and ESD. We conducted experiments for different percentages of examples with missing features, ranging from 10% to 70%. For each of these percentages, 30 similar trials were run, in which examples were selected at random to have some of its entries erased. Once an example has been selected, up to a third of its features are selected randomly to be deleted .In each of the trials, the GMM parameters were computed using Expectation Maximization and the number of components was chosen through model selection using the Bayesian Information Criterion. Up to 10 components were considered for each GMM. Tables 5 and 6 report the RMSE and the relative success rate performances of EED, ESD and CMI when applied to real-world data. Again, statistical significance analysis was performed following the guidelines given in Section 5.2.

With regard to the RMSE performance, we observe that EED outperforms CMI and ESD in all scenarios, i.e., with small and large amount of missing data. As expected, the performance gap between EED and ESD is smaller than the gap between EED and CMI. The hypothesis test indicates significant difference between EED and the other methods. Considering the relative success rate, EED reports a steady superior performance regardless the amount of missing data. Again, the hypothesis test indicates statistical significance of the achieved results.

5.4. MLM for datasets with missing values

In the last experiment, we evaluate the performance of EED-MLM on the same real-world datasets used in Section 5.3. For this purpose, we compare EED-MLM to (i) MLMs trained using CMI to directly impute missing input data entries; (ii) MLMs trained using the ESD to estimate $\mathbf{D}_x$ [26]; and (iii) MLMs trained by discarding feature vectors with unobserved components (Listwise Deletion,

Table 5
RMSE performances on real world datasets. The symbols ✓ and ✗ indicate the result of the hypothesis test (✓ fail to reject, and ✗ reject).

Dataset	**Method**	**RMSE**						
		Missing Data (%)						
		10	20	30	40	50	60	70
AP	ESD	13.50 ✗	20.57 ✗	25.18 ✗	28.05 ✗	32.81 ✗	37.38 ✗	38.40 ✗
	CMI	13.63 ✗	22.17 ✗	26.17 ✗	28.78 ✗	34.46 ✗	39.20 ✗	39.10 ✗
	EED	13.38	20.50	24.97	27.79	32.60	37.19	38.11
AM	ESD	37.64 ✗	51.55 ✗	63.73 ✗	75.30 ✗	84.93 ✗	94.07 ✗	95.46 ✗
	CMI	40.34 ✗	55.15 ✗	69.61 ✗	82.87 ✗	94.27 ✗	105.79 ✗	106.91 ✗
	EED	37.33	50.94	63.23	74.89	84.41	93.71	94.67
BH	ESD	63.06 ✗	112.14 ✗	131.11 ✗	136.98 ✗	141.21 ✗	170.83 ✗	167.73 ✗
	CMI	67.16 ✗	118.67 ✗	140.90 ✗	147.59 ✗	151.86 ✗	187.82 ✗	183.06 ✗
	EED	62.84	111.97	130.90	136.54	140.52	170.40	166.91
CC	ESD	116.43 ✗	165.79 ✗	183.31 ✗	223.86 ✗	254.04 ✗	309.23 ✗	311.74 ✗
	CMI	124.49 ✗	176.77 ✗	196.42 ✗	243.82 ✗	276.76 ✗	339.61 ✗	343.68 ✗
	EED	115.46	164.42	181.28	222.09	251.86	307.57	309.48
SV	ESD	24.45 ✗	30.56 ✗	37.07 ✗	45.35 ✗	48.75 ✗	55.09 ✗	57.95 ✗
	CMI	27.57 ✗	34.36 ✗	42.21 ✗	53.16 ✗	57.38 ✗	67.09 ✗	70.04 ✗
	EED	24.25	30.18	36.61	44.96	47.96	54.57	57.08

Table 6
Relative sucess rate of EED vs CMI and ESD on real-world datasets.

Dataset	**Method**	**Relative success rate**						
		Missing Data (%)						
		10	20	30	40	50	60	70
AP	CMI	0.96 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	ESD	0.80 ✗	0.86 ✗	0.90 ✗	0.86 ✗	0.86 ✗	0.90 ✗	0.90 ✗
AM	CMI	0.86 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	ESD	0.83 ✗	0.93 ✗	0.73 ✗	0.83 ✗	0.86 ✗	0.83 ✗	0.83 ✗
BH	CMI	0.86 ✗	1.00 ✗	0.96 ✗	0.96 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	ESD	0.70 ✗	0.70 ✗	0.63 ✗	0.73 ✗	0.86 ✗	0.93 ✗	0.83 ✗
CC	CMI	0.96 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	ESD	0.80 ✗	1.00 ✗	1.00 ✗	1.00 ✗	0.96 ✗	1.00 ✗	1.00 ✗
SV	CMI	0.96 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	ESD	0.80 ✗	0.83 ✗	0.86 ✗	0.76 ✗	0.93 ✗	0.93 ✗	1.00 ✗

Table 7
RMSE results of MLMs trained using ESD, CMI, LSD and EED. The symbols ✓ and ✗ indicate the result of the hypothesis test (✓ fail to reject, and ✗ reject).

Dataset	**Method**	**RMSE**						
		Missing Data (%)						
		10	20	30	40	50	60	70
AP	LSD	1862.76 ✗	1853.25 ✗	1946.07 ✗	2045.06 ✗	2221.65 ✗	2268.52 ✗	2417.23 ✗
	CMI	1751.54 ✓	1728.64 ✓	1760.52 ✗	1859.19 ✓	1958.69 ✗	2092.73 ✓	2171.96 ✗
	ESD	1757.66 ✓	1727.79 ✓	1757.54 ✓	1859.75 ✓	1944.30 ✓	2091.69 ✓	2165.50 ✓
	EED	1757.14	1727.86	1757.04	1859.43	1944.20	2091.56	2165.23
AM	LSD	2.35 ✗	2.52 ✗	2.58 ✗	2.67 ✗	2.77 ✗	2.83 ✗	3.14 ✗
	CMI	2.27 ✗	2.41 ✗	2.47 ✗	2.55 ✗	2.55 ✗	2.63 ✓	2.78 ✗
	ESD	2.26 ✓	2.39 ✓	2.43 ✓	2.52 ✓	2.53 ✗	2.62 ✓	2.75 ✓
	EED	2.26	2.39	2.43	2.51	2.52	2.62	2.75
BH	LSD	3.10 ✗	3.70 ✗	3.73 ✗	4.00 ✗	4.11 ✗	4.52 ✗	4.63 ✗
	CMI	2.95 ✗	3.34 ✗	3.35 ✗	3.55 ✗	3.65 ✗	3.94 ✗	3.98 ✓
	ESD	2.96 ✓	3.33 ✗	3.33 ✗	3.54 ✗	3.63 ✗	3.97 ✗	3.99 ✗
	EED	2.96	3.32	3.33	3.54	3.63	3.96	3.98
CC	LSD	5.28 ✗	5.90 ✗	6.04 ✗	6.38 ✗	6.75 ✗	7.14 ✗	7.86 ✗
	CMI	5.07 ✗	5.55 ✓	5.80 ✗	6.09 ✗	6.34 ✗	6.60 ✗	6.97 ✗
	ESD	5.01 ✓	5.52 ✓	5.72 ✓	5.99 ✗	6.25 ✗	6.56 ✗	6.89 ✗
	EED	5.01	5.51	5.72	5.97	6.22	6.53	6.84
SV	LSD	0.58 ✓	0.76 ✗	0.76 ✗	0.75 ✗	0.83 ✗	0.85 ✗	0.89 ✗
	CMI	0.58 ✗	0.70 ✗	0.71 ✗	0.70 ✗	0.75 ✗	0.78 ✗	0.80 ✗
	ESD	0.57 ✓	0.68 ✗	0.70 ✗	0.68 ✓	0.73 ✓	0.75 ✓	0.78 ✓
	EED	0.57	0.67	0.69	0.68	0.73	0.75	0.78

Table 8
Relative success rate of MLM-EED vs. MLMs trained using ESD, LSD and CMI. The symbols ✓ and ✗ indicate the result of the hypothesis test (✓ fail to reject, and ✗ reject).

Dataset	Method	Relative Success Rate						
		Missing Data (%)						
		10	20	30	40	50	60	70
AP	LSD	0.73 ✗	0.87 ✗	0.83 ✗	0.93 ✗	0.93 ✗	0.87 ✗	0.93 ✗
	CMI	0.40 ✓	0.63 ✗	0.73 ✗	0.40 ✓	0.77 ✗	0.60 ✓	0.80 ✗
	ESD	0.63 ✗	0.50 ✓	0.63 ✗	0.53 ✓	0.63 ✗	0.53 ✓	0.60 ✓
AM	LSD	0.90 ✗	0.93 ✗	0.93 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	CMI	0.60 ✓	0.60 ✓	0.80 ✗	0.77 ✗	0.77 ✗	0.70 ✗	0.67 ✗
	ESD	0.40 ✓	0.60 ✓	0.43 ✓	0.50 ✓	0.83 ✗	0.67 ✗	0.63 ✗
BH	LSD	0.77 ✗	0.87 ✗	0.87 ✗	1.00 ✗	1.00 ✗	0.90 ✗	1.00 ✗
	CMI	0.33 ✗	0.63 ✗	0.87 ✗	0.77 ✗	0.77 ✗	0.13 ✗	0.60 ✗
	ESD	0.63 ✗	0.93 ✗	0.67 ✗	0.80 ✗	0.70 ✗	1.00 ✗	0.87 ✗
CC	LSD	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗	1.00 ✗
	CMI	1.00 ✗	0.63 ✗	0.83 ✗	0.80 ✗	0.60 ✓	0.70 ✗	0.97 ✗
	ESD	0.57 ✓	0.63 ✗	0.57 ✓	0.80 ✗	0.83 ✗	0.87 ✗	0.90 ✗
SV	LSD	0.63 ✗	0.97 ✗	0.87 ✗	0.73 ✗	0.93 ✗	0.97 ✗	0.93 ✗
	CMI	0.70 ✗	0.87 ✗	0.87 ✗	0.80 ✗	0.80 ✗	0.87 ✗	0.57 ✓
	ESD	0.40 ✓	0.90 ✗	0.93 ✗	0.53 ✓	0.50 ✓	0.57 ✓	0.60 ✓

LSD). As before, 30 independent runs of experiments were held. To make a fair comparison, for each trial, all MLMs were trained using the same reference points, which were chosen by drawing half of the fully observed training examples at random. Performance was measured in terms of RMSE and relative success rate and the results are shown in Tables 7 and 8.

As expected, the listwise deletion method had the worst performance overall, being outperformed by EED in all datasets and amounts of missing data. Analyzing the results of CMI and ESD we note that, in general, when the number of missing instances is low, its relative success rates are near 50% and no significant difference is observed in terms of RMSE. These results indicate that the quality of the distance estimation does not have a significant impact on the MLM performance when the number of missing data is low. However, in most cases, as the number of missing data increases so does the relative success rate, and it starts to produce statistically significant results in terms of RMSE. It is interesting to note that, in some situations, EED is outperformed by CMI even when the number of instances with missing values is high. A possible reason is that there is no guarantee that improving the estimates of the Euclidean distance necessarily leads to improvements on the performance of the MLM even though it might be the general case.

6. Conclusion

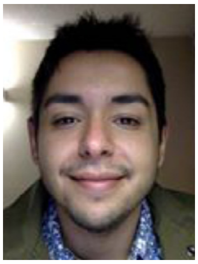
In this paper, we presented an algorithm to estimate the pairwise Euclidean distance between vectors with missing data. In the proposed method, named Expected Euclidean Distance (EED), the distance between vectors is said to follow a Nakagami distribution whose parameters can be calculated using the first four moments of the dataset distribution. In EED, a GMM is used to model the distribution of the dataset.

The accuracy of EED in distance estimation is verified using experiments on synthetic and real-world datasets with varying quantities of entries with missing components. On the basis of our experiments, we can state that EED is a valid alternative for distance estimation in incomplete data, since it outperformed other commonly used methods. In the last experiment, we also verified that the improved distance estimates impacted the results obtained by the MLM for datasets with missing values.

References

[1] E. Eirola, A. Lendasse, V. Vandewalle, C. Biernacki, Mixture of gaussians for distance estimation with missing data, Neurocomputing 131 (2014) 32–42.
[2] E. Eirola, G. Doquire, M. Verleysen, A. Lendasse, Distance estimation in numerical data sets with missing values, Inform. Sci. 240 (2013) 115–128.
[3] P. Kang, Locally linear reconstruction based missing value imputation for supervised learning, Neurocomputing 118 (2013) 65–78.
[4] F. Lobato, C. Sales, I. Araujo, V. Tadaiesky, L. Dias, L. Ramos, A. Santana, Multi-objective genetic algorithm for missing data imputation, Pattern Recog. Lett. 68, Part 1 (2015) 126–131.
[5] M. Aste, M. Boninsegna, A. Freno, E. Trentin, Techniques for dealing with incomplete data: a tutorial and survey, Pattern Anal. Appl. 18 (1) (2015) 1–29.
[6] I.A. Gheyas, L.S. Smith, A neural network-based framework for the reconstruction of incomplete data sets, Neurocomputing 73 (16–18) (2010) 3039–3065. 10th Brazilian Symposium on Neural Networks (SBRN2008)
[7] E. Acuña, C. Rodriguez, The Treatment of Missing Values and its Effect on Classifier Accuracy, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 639–647.
[8] D. Sovilj, E. Eirola, Y. Miche, K.-M. Björk, R. Nian, A. Akusok, A. Lendasse, Extreme learning machine for missing data using multiple imputations, Neurocomputing 174, Part A (2016) 220–231.
[9] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27.
[10] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, Calif., 1967, pp. 281–297.
[11] T. Kohonen, Self-organization and Associative Memory: 3rd Edition, Springer-Verlag New York, Inc., New York, NY, USA, 1989.
[12] M. Nakagami, The m-distribution–A general formula of intensity distribution of rapid fading, in: W.C. Hoffmann (Ed.), Statistical Methods in Radio Wave Propagation, Elmsford, NY, 1960.
[13] X.-L. Meng, D.B. Rubin, Maximum likelihood estimation via the ECM algorithm: a general framework, Biometrika 80 (2) (1993) 267–278.
[14] A.H. Souza Júnior, F. Corona, G.A. Barreto, Y. Miche, A. Lendasse, Minimal learning machine: a novel supervised distance-based approach for regression and classification, Neurocomputing 164 (2015) 34–44.
[15] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley-Interscience, 2002.
[16] Y. Ding, J.S. Simonoff, An investigation of missing data methods for classification trees applied to binary response data, J. Mach. Learn. Res. 11 (2010) 131–170.
[17] J.K. Dixon, Pattern recognition with partly missing data, IEEE Trans. Syst. Man Cybern. 9 (10) (1979) 617–621.
[18] C. Roberts, S. Geisser, A necessary and sufficient condition for the square of a random variable to be gamma, Biometrika Trust 53 (1/2) (1966) 275–278.
[19] A. Azzalini, A class of distributions which includes the normal ones, Scand. J. Stat. 12 (2) (1985) 171–178.
[20] K.S. Johnson N., N. Balakrishnan, Continuous Univariate Distributions, Wiley, 1995.
[21] M.G. Genton, Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, Chapman and Hall/CRC, 2004.
[22] S. Covo, A. Elalouf, A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions, Electron. J. Statist. 8 (1) (2014) 894–926.

[23] G.C. Alexandropoulos, A. Papadogiannis, K. Berberidis, Performance analysis of cooperative networks with relay selection over nakagami- m fading channels, IEEE Signal Process. Lett. 17 (5) (2010) 441–444.
[24] L. Hunt, M. Jorgensen, Mixture model clustering for mixed data with missing information, Comput. Stat. Data Anal. 41 (3–4) (2003) 429–440.
[25] M. Lichman, UCI Machine Learning Repository, 2013.
[26] D.P.P. Mesquita, J.P.P. Gomes, A.H. Souza Jr, A minimal learning machine for datasets with missing values, in: Proceedings of the 22nd International Conference on Neural Information Processing, ICONIP 2015, Istanbul, Turkey, November 9–12, 2015Part I, Springer International Publishing, 2015, pp. 565–572.



Diego Parente P. Mesquita a bachelor's degree on Computer Science from Universidade Federal do Ceará (UFC, 2015) and is currently pursuing a master's degree in Computer Science at the same university.

João Paulo Pordeus Gomes holds a bachelor's degree on Electrical Engineering from Universidade Federal do Ceará (UFC, 2004), Brazil, master's (2006) degree on aeronautical Engineering and doctorate's (2011) degree in electronic engineering from Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, SP, Brazil. Dr. Gomes worked for EMBRAER S.A. between 2006 and 2013, as a Technology Development Engineer focusing on fault monitoring applications on aeronautical systems. He is currently an Assistant Professor at UFC

Amauri Holanda de Souza Junior holds a bachelor's degree on Telematics from Federal Institute of Ceará (IFCE, 2007), Brazil, master's (2006) degree and doctorate's (2014) degree in Teleinformatics Engineering from Universidade Federal do Ceará (UFC),Fortaleza, CE, Brazil. He is currently an Assistant Professor at IFCE.

Juvêncio Santos Nobre holds a bachelor's degree on Statistics from Universidade Federal do Ceará (UFC, 2002), Brazil, master's (2004) degree and doctorate's (2017) degree in Statistics from Universidade de São Paulo (USP), São Paulo, SP, Brazil. He is currently an Assistant Professor at UFC.