LINEAR REGRESSION BASED ON CORRENTROPY FOR SENSOR CALIBRATION

Joilson B.Á Régo*, Aluisio I.R Fontes†, Adrião D.D Neto‡, Luiz F.Q. Silveira§, Allan de M. Martins¶

*Instituto de Computação. Universidade de Alagoas, 57072900, Maceió, AL, Brasil
†Departamento de Engenharia de Computação e Automação, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brasil.
‡Departamento de Engenharia de Computação e Automação, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brasil.
§Departamento de Engenharia de Computação e Automação, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brasil.

Emails: jotarego@gmail.com, aluisio@dca.ufrn.br, adriao@dca.ufrn.br, lfelipe@dca.ufrn.br, allan@dee.ufrn.br

Abstract— Correntropy is a measure of statistical similarity between two random variables. This paper proposes an extension of correntropy for $N$ random variables, and sets a new linear regression method from this extension. Experiments carried out in the presence of impulse noise demonstrated the efficiency of the proposed method, taking as reference, the sensor calibration problem.

Keywords— Correntropy, Linear Regression, Sensor Calibration.

Resumo— A correntropia é uma medida de similaridade estatística entre duas variáveis aleatórias. Este trabalho propõe uma extensão da correntropia para $N$ variáveis aleatórias, e define um novo método de regressão linear a partir desta extensão. Experimentos realizados na presença de ruído impulsivo demonstraram a eficiência da ferramenta proposta, tomando-se como referência, o problema de calibragem de sensores.

Palavras-chave— Correntropia, Regressão Linear, Calibragem de Sensores.

1 Introduction

The concept of similarity is fundamental in statistical measures when the objective is to compare two random variables. Statistical second-order in the form of correlation and mean square error (MSE) are probably the most widely methods used to quantify how similar are two random variables (Singh and Principe, 2010). The troubleshooting using second order statistics depend on the presence of gaussianity and linearity in the data, so that the existence of impulse noise (outliers) in the measurements makes the estimation error increases significantly.

In recent years, data fitting approaches using mean square error has been improved with the inclusion of criteria based on information theory, setting a new field called Information Theoretic Learning (ITL). In (Santamaria et al., 2006) extended the concept of correlation to a new similarity measure named correntropy. The correntropy is a robust similarity measure obtained for two random variables, which uses a symmetric and positive definite kernel function.

Correntropy is directly related to the probability of two random variables are similar in a sample space controlled by a parameter called kernel size. The adjustment of the kernel size gives an efficient mechanism to eliminate outliers. However, both the MSE and correntropy are similarity measure defined for only two random variables. Furthermore, in the context of correntropy, the kernel size should be adjusted according to the application (Liu et al., 2006).

This paper proposes a generalization of correntropy for $N$ random variables, even with different dimensions, using statistical concepts. This new measure was evaluated in a sensor calibration application. The results show good efficiency and lower dependence between the application and the parameter kernel size.

The remainder of the paper is organized as follows. Section 2 presents concepts related to correntropy as well as theoretical tools used in the development of the proposed method. Section 3 presents the considered problem, and results obtained from the experimental analysis of the linear regression method based on correntropy. Finally, Section 4 presents some final remarks on the proposed approach.

2 Correntropy

Correntropy is a generalization of the correlation measure between random signals defined as (Santamaria et al., 2006):
In this work, \( G(\ldots) \) is a Gaussian kernel given as

\[
G_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(X, Y)^2}{2\sigma^2} \right].
\]

where \( \sigma \) is the variance defined as the kernel size.

Figure 1 shows diagrammatically that correntropy provides the probability density of the event \( p(x = y) \). This understanding is interesting, because indeed we are able to quantify the probability of two events being equal (Liu et al., 2006).

Nowadays, correntropy has been successfully used in a wide variety applications where the signals are nongaussian or nonlinear e.g., automatic modulation classification (Fontes, Martins, Silvera and Principe, 2014), classification systems of pathological voices (Fontes, Souza, Neto, Martins and Silvera, 2014), principal component analysis (He et al., 2011), and adaptive filter (Singh and Principe, 2009).

### 3 Correntropy for Regression

In the proposed problem, the goal is to obtain a conversion constant that relates two values linearly as

\[
T = aV
\]

This problem is, at first view, very trivial and actually consists in a one variable regression problem. However, as we will see in the results section, in the presence of impulse noise or when the measure has outliers, Correntropy can be used to naturally weight off those outliers (Santamaria et al., 2006). There is still a problem to be solved when using Correntropy regarding the roughness of cost function. In order to avoid outliers, one must pick, in general, a small kernel size that will model the error probability of the parameter well. Due to the finite the limited number of points available in practical problems, the Parzen’s estimate for the probability is noisy. That problem can cause the optimization of the parameters very difficult.

To overcome this difficulty, we will propose a generalization for Correntropy in the light of the interpretation of section 2. As it was showed, Correntropy is a measure of a Parzen’s estimate for the probability of both sides of 1 being equal. In this problem, one would find the constant that maximizes that probability. When we have just a limited number of data points (pairs of values for \( T \) and \( V \)), that estimate can become noisy and imprecise. Moreover, one must pick a suitable kernel size.

One of the possible generalizations is to consider several random variables being equal one another (not just two). In this case, one must extend the idea of integration along the 45 degree line and integrate along the line \( x_1 = x_2 = x_3 = \ldots = x_n \). Using this approach, we need to reformulate the problem in the following manner. We have variables and measurements for each. We want to maximize the following probability

\[
av_1^i = av_2^i = av_3^i = \ldots = av_L^i = T^i
\]
and use the logarithm of the quantities (hence finding the logarithm of the constant $T^i$). Now, the expression in 6 can be rewritten as

$$a' + \lambda_1^i = t^i = a' + \lambda_2^i = t^i = ... = a' + \lambda_L^i = t^i \quad (7)$$

where $a' = \log(a)$, $\lambda_1^i = \log(v_1^i)$, and $t^i = \log(T^i)$.

Using a parzen’s estimate for the probability and assuming the use of the logarithms, we can compute that probability Equation 7 as

$$p = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2\pi \sigma^2} e^{-\frac{(x - k)^2}{2\sigma^2}} \sqrt{\frac{C}{4}} \quad (8)$$

where

$$C = 2(t^i)^2 + \frac{2}{L} \sum_{j=1}^{L} (a' + \lambda_j^i)^2 - \left(t^i + a' + \frac{1}{L} \sum_{j=1}^{L} \lambda_j^i \right)^2 \quad (9)$$

Now, maximizing Equation 7 for $a$ will find the best value to use in the relation showed in 5.

### 4 Correntropy in Sensors Calibration

In order to show the performance of the new expression for computing the constant $a$, we will use as an example, a problem of calibrating a temperature sensor. The idea is to find the best conversion constant (in °C per volts) that translates a measured voltage $V$ into a temperature $T$. The relation is the same as presented in Equation 5. In this example we consider $L$ sensors and perform $N$ measurements each. We also have $N$ known temperatures so, for each of the $N$ measurements at each $L$ sensor, we compare against the corresponding temperature. For the purpose of this paper, we will use only 5 measurements and for only 5 sensors. Moreover, it is possible to notice two outliers that have a huge impact no Mean Square Error (MSE) estimation, as we will see in the results. Normally this is a difficult problem due to the lack of more data, but, as we can see, using Correntropy and further, is generalization, we can estimate the constant $a$ with a good precision. In Figure 2 shows a plot of one example of data points used in this paper.

Equation 8 has been evaluated according to Monte Carlo’s method. For each experiment a minimum number of 200 trials were used by computational simulation carried out in MATLAB. The constant $a$ was adjusted to 30 and kernel size to 0.5. The proposed method has been evaluated and compared with MSE and classical correntropy defined by Equation 4.

Figure 3 shows that the proposed method has lower sensitivity to variations of the kernel size, in contrast with the conventional correntropy, which is highly dependent of this parameter.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>15.05</td>
<td>4.42</td>
</tr>
<tr>
<td>Correntropy</td>
<td>8.36</td>
<td>33.79</td>
</tr>
<tr>
<td>This work</td>
<td>1.68</td>
<td>1.42</td>
</tr>
</tbody>
</table>

We can observe a significant difference between the performances of the algorithms. It is
clear that the proposed method is both smooth as well as robust to the presence of noise. In addition, the method is insensitive to the kernel size, like MSE, and is able to filter Gaussian noise.

Analyzing figures 3, 4 and 5, we conclude that the proposed method is a Minimum Variance Estimator, which encompasses two main advantages. The smoothness, like the MSE, and the robustness to outliers, similarly to the conventional Correntropy.

5 Conclusion

This paper develops a new algorithm that extends the application of Correntropy to linear estimation. The new extensions compute Correntropy in any dimension rather than only comparing two scalar variables as in the original definition. Due to this fact each extension can lead to a different interpretation in terms of probability. The estimators for the probabilities are all based on the Parzen estimation of the joint pdf of the data, therefore the kernel size of the estimator has to be chosen. The experiments showed that for the linear estimation application the choice of the kernel size does not interfere too much in the final results, leading to a very good and practical algorithm to be used in real situations.

Acknowledgment

The authors acknowledge CNPQ (process Number 486135/2013-6) for the financial support.

References


