Electric Power Systems Research 157 (2018) 261–271

Contents lists available at ScienceDirect

Electric Power Systems Research

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

State estimation pre-filtering with overlapping tiling of autoencoders☆

Marco A.M. Saran [a,b,*], Vladimiro Miranda [a,b]

[a] INESC TEC, INESC Technology and Science, Porto, Portugal[1]
[b] FEUP, Faculty of Engineering, University of Porto, Portugal[2]

ARTICLE INFO

Article history:
Received 21 June 2017
Received in revised form
28 November 2017
Accepted 26 December 2017

Keywords:
Autoencoder
Bad data
Correntropy
Data fusion
Gross error
Neural network
State estimation

ABSTRACT

This paper presents a new concept for an approach to deal with measurements contaminated with gross errors, prior to power system state estimation. Instead of a simple filtering operation, the new procedure develops a screen-and-repair process, going through the phases of detection, identification and correction of multiple gross errors.

The method is based on the definition of the coverage of the measurement set by a tiling scheme of 3-overlapping autoencoders, trained with denoising techniques and correntropy, that produce an ensemble-like set of three proposals for each measurement. These proposals are then subject to a process of fusion to produce a vector of proposed/corrected measurements, and two fusion methods are compared, with advantage to the Parzen Windows method. The original measurement vector can then be recognized as clean or diagnosed with possible gross errors, together with corrections that remove these errors. The repaired vectors can then serve as input to classical state estimation procedures, as only a small noise remains. A test case illustrates the effectiveness of the technique, which could deal with four simultaneous gross errors and achieve a result close to full recognition and correction of the errors.

1. Introduction

A control center, either in transmission or distribution, cannot function without some kind of state estimation. The huge transformation being witnessed at the distribution level, with the emergence of distributed (and uncontrolled) generation, has just reinforced the need of system operators (TSO and DSO) to be able to monitor, at all times, the state of the networks, However, especially when closer to the distribution level, but also with PMU measurements due to wrong or absent time-tagging, gross errors tend to appear in the measurement sets observed at any moment. It may be stated that handling this problem properly is one central concern in the architecture of a modern SCADA/EMS-DMS system. Handling one gross error has had, in the past, some success with classical techniques working on residuals (the difference between measurements and estimated values), but the same cannot be said about handling multiple errors — and with a widespread monitoring including distribution, the ability to handle multiple gross errors becomes a necessity.

The classical and most known methods to identify gross errors are: the Chi-squares Test, the Largest Normalized Residual Test and the Hypothesis Testing Identification [1,2]. These methods are applied only after each estimation iteration and are centered on the residuals. The obvious conceptual flaw is that they rely on post-processing and depart from contaminated results. As a consequence, they exhibit some failure rate in detecting bad data. Moreover, many of these methods depend on different assumptions regarding the system and the errors characteristics. Some of these assumptions are controversial and generate debate, mainly about the gaussianity [3] and the independence [4,5] of the errors.

An additional difficulty derives from the fact that, in many of the conventional error handling methods, there is no provision to recover an assumed erroneous measurement, and this is simply removed from the data set [1,2]. This reduces the redundancy of the input measurement set, discards information that could be useful and, in severe cases, hampers the observability of the network.

☆ The authors acknowledge the support of ERDF — European Regional Development Fund (COMPETE 2020 Programme) and of FCT — Fundação para a Ciência e a Tecnologia (project POCI-01-0145-FEDER-016731 INFUSE). M. Saran further acknowledges the support of FCT through Ph.D. Scholarship SFRH/BD/68671/2010, Portugal.

* Corresponding author at: INESC TEC, INESC Technology and Science, Porto, Portugal.

E-mail addresses: marco@msaran.com (M.A.M. Saran), vmiranda@inesctec.pt (V. Miranda).

[1] http://www.inesctec.pt.
[2] http://www.fe.up.pt.

https://doi.org/10.1016/j.epsr.2017.12.026
0378-7796/

Meanwhile, in a completely alternative path, the work reported in Ref. [6] made a robust demonstration that neural networks with special architectures, denoted autoencoders, are tools that, properly handled, can learn the supporting manifold of system state patterns — and then they can be used to correct measurement vectors that either have components with gross errors or are corrupted and exhibit missing signals in some of the components.

There was also a lesson learned: that a very large autoencoder, representing at its input the whole set of measurements, becomes a cumbersome artifact to be trained. But, at the same time, the need for such a huge neural network is also questioned, and a distinct conceptual model was proposed (also in Ref. [7]): a mosaic of adjacent local areas representing the network, each cell being observed by an autoencoder. The advantage of this scheme would derive from the small scale of each neural network to be tuned, from the fact that steady state causes have only visible local effects and from the easy adaptation of this concept to system changes, because only local retraining would be necessary in case of structural changes of the network.

This paper is devoted to exploring the potential of autoencoders to act as pre-filters to the measurement vector, and thus perform the three necessary functions of an ideal system: detection, identification and quantification. Here is the definition of these terms:

- Detection: the ability to signal out that a data set contains bad data
- Identification: the ability to pinpoint which measurement is corrupt
- Quantification (or repair, or correction): the ability to estimate the amount needed to be added to the identified corrupt measurement to bring it to a value coherent with the physical system under observation (in the power systems case, this should be the Kirchhoff laws).

If these three functions are successfully performed, the data set remains intact (albeit corrected), no observability is lost and classical state estimation methods can even take on from there, if required. No post-processing will be needed.

There are some guidelines to be followed, if a successful method is devised: it must be fast enough to be applied in real-time; it should be non-parametric, i.e. independent of the network or measurement parameters or any error assumption; it should deal with the possibility of having multiple errors originating from the same cause and not being independent; and it should be applied in a preprocessing fashion. A new form of efficient data pre-filtering would result.

The concept described in this paper has the autoencoder as a common trait with the work in Ref. [6]. However, apart from this, it displays distinct options and choices, makes use of a different mix of algorithms, based on computational intelligence with elements of information theoretic learning, machine learning and data fusion, and proposes a different arrangement for the mosaic of autoencoders observing the network — instead of a tessellation, an overlapping tiling in now used, taking advantage of having the same node monitored by more than one autoencoder. The results of such a new concept are impressive, in handling multiple gross errors.

2. Fundamental concepts

As the technique described in this paper involves a set of concepts, the following sections will make an abridged reference to the most important.

2.1. Parzen Windows

The Parzen Windows technique [8] is a non-parametric method to infer an approximate pdf – probability density function $\hat{p}(y)$, directly from discrete data, expressed as:

$$\hat{p}(y) = \frac{1}{|T|}\sum_{y_j \in T} K_j\left(y - y_j\right) \tag{1}$$

where K_j represent kernel functions centered on the T points in a sample of instances Y.

The most used kernel function is the Gaussian kernel, as it is smooth and hence the density function also varies smoothly, and is defined as:

$$G_\sigma(x) = \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \tag{2}$$

where σ is denoted as the window size.

2.2. Correntropy

The correntropy V_σ for two discrete random variables X and Y can be written as:

$$V_\sigma(X, Y) = \frac{1}{n}\sum_{i=1}^{n} G_\sigma\left(x_i - y_i, \sigma^2 I\right) \tag{3}$$

where $k_\sigma(x_i - y_i, \sigma^2 I)$ is a Gaussian kernel with size σ. In Ref. [9], the concept of *correntropy* is extensively discussed as a localized similarity measure. Correntropy has a relation with information with information entropy and a set of interesting properties [10–12]. Among these, it is interesting to note that correntropy corresponds to the probability of having $X = Y$ — it is the integral of the marginal distribution of the joint (X,Y) distribution, obtained for the condition $x = y$. If we write $\varepsilon_i = x_i - y_i$, we can define the correntropy of an error distribution.

In the training of mappers, the MCC — Maximum Correntropy Criterion, is defined as:

$$\max\ V_\sigma(\varepsilon) = \frac{1}{n}\sum_{i=1}^{n} G_\sigma\left(\varepsilon_i, \sigma^2 I\right) \tag{4}$$

where $\sigma^2 I$ represents the covariance matrix (assumed with independent and equal variances in all dimensions). The kernel size acts as an observation window and provides an automatic mechanism to eliminate the effect of outliers, being intrinsically different from the conventional techniques that use some sort of threshold.

2.3. Quadratic Mutual Information

The Quadratic Mutual Information (QMI) can be interpreted as a similarity criterion. It is based on Renyi's definition of Entropy, and bears resemblance to Shannon's Mutual Information [13,14], which is a standard measure of statistical dependence of random variables. The QMI uses Euclidean distance instead of the Kullback–Leibler divergence and has the advantage of being able to be easily integrated using Parzen Window method, if needed. It is, for some purposes, a simple way to estimate mutual information, with a significant improvement in computing time and, additionally, it allows to achieve an efficient non-parametric estimation requiring no prior assumptions.

The Quadratic Mutual Information was proposed as the Cauchy-Schwarz divergence between the joint and the product of the