Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 11 (2017) 2018 – 2024

Procedia MANUFACTURING

27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy

Anomaly Detection Approaches for Semiconductor Manufacturing

Gian Antonio Susto*, Matteo Terzi, Alessandro Beghi

Department of Information Engineering, University of Padova, via Giovanni Gradenigo 6/B, 35131 Padova, Italy
Human Inspired Technology Centre, University of Padova, via Luzzatti 4, 35121 Padova, Italy

Abstract

Smart production monitoring is a crucial activity in advanced manufacturing for quality, control and maintenance purposes. Advanced Monitoring Systems aim to detect anomalies and trends; anomalies are data patterns that have different data characteristics from normal instances, while trends are tendencies of production to move in a particular direction over time. In this work, we compare state-of-the-art ML approaches (ABOD, LOF, onlinePCA and osPCA) to detect outliers and events in high-dimensional monitoring problems. The compared anomaly detection strategies have been tested on a real industrial dataset related to a Semiconductor Manufacturing Etching process.

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing

Keywords: Anomaly Detection, Advanced Monitoring Systems, Data Mining, Etching, Fault-diagnosis, Machine Learning, Optical Emission Spectroscopy, Semiconductor Manufacturing

1. Introduction

Smart production monitoring is a crucial activity in advanced manufacturing for quality [1], control [2, 3] and maintenance purposes [4]. Advanced Monitoring Systems (AMSs) aim at detecting anomalies and trends; anomalies

* Corresponding author. Tel.: +39 049-827-7760.
E-mail address: gianantonio.susto@dei.unipd.it

Peer-review under responsibility of the scientific committee of the 27th International Conference on Flexible Automation and Intelligent Manufacturing
doi:10.1016/j.promfg.2017.07.353

are data patterns that have different data characteristics from normal instances [5], while trends are tendencies of production to move in a particular direction over time.

Instruments to implement efficient AMSs are provided by Machine Learning (ML). ML approaches have proliferated in recent years Advanced Process Control (APC) solutions for Semiconductor Manufacturing [6], thanks to the algorithmic advancements in the field and the increased computational and storage capabilities in the IT architecture of the Fabs; ML-based approaches have been used for Virtual Metrology, Predictive Maintenance and Fault Detection applications. In this work, we compare state-of-the-art ML approaches to detect outliers and events in high-dimensional monitoring problems.

The compared anomaly detection strategies have been tested on a real industrial dataset related to a Semiconductor Manufacturing Etching process.

The contributions of the paper are: (i) compare state-of-the-art anomaly detection techniques, (ii) apply such techniques in a real industrial Semiconductor Manufacturing applications and (iii), to the best of our knowledge, applying such techniques for the first time in Semiconductor Manufacturing.

The rest of the paper is organized as follows: Section 2 is dedicated to present and compare the anomaly detection methodologies employed in this work; in Section 3 the Semiconductor Manufacturing case study is illustrated and the Experimental results are provided. Finally, in Section 4 the concluding remarks are provided.

A list of the notation employed in this work is reported here:

Nomenclature

N	Observations available in the dataset
p	Number of features monitored
$q \in [1, \dots, p]$	A generic feature
$X = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^p$	Input matrix
x_{new}	New observation, to be monitored/classified
s	Anomaly Score (AS)
τ	Threshold on AS
m	[osPCA-onlinePCA] Iterations of the Power Method
k	[ABOD] Number of considered neighbours
$d_k(x)$	[LOF] Distance of $x \in \mathbb{R}^p$ from its k-nearest neighbor
$r_k(y,x) = max\{d_k(x), d(y,x)\}$	[LOF] Reachability distance
lrd_k	[LOF] Local Reachability distance
K	Monte Carlo Simulations

2. Anomaly Detection Methodologies

In this Section, a list of the compared Anomaly Detection methodologies is presented. Each of the listed techniques define an *Anomaly score (AS)* s: a quantitative index that defines the 'outlierness' degree of an observations. Automatic Monitoring policies are generally based on triggering a reaction/intervention if the AS for a new observation is above a predefined threshold τ.

2.1. osPCA and onlinePCA

OsPCA [7] is an anomaly detection method based on the analysis Principal Component Analysis (PCA) [8]. Given a matrix $X \in \mathbb{R}^{\{N \times p\}}$, PCA employs an orthogonal transformation in order to find a new orthogonal basis (columns are linearly independent) whose elements are called principal components (PCs). Informally, PCs are the most informative directions (with highest explained variance) decreasingly ordered (first PC is the most informative).

The idea underlying osPCA is that principal directions represent "normal attributes" and outliers change the direction of principal components: in order to reduce complexity only the first PC is considered. However, outliers occur with a very lower frequency w.r.t. normal attributes and thus the change in direction of PCs may not be appreciated. In order to address this issue, in osPCA an oversampling of the tested attribute x_{new} which is replicated $\tilde{n}$ times. If x_{new} is an outlier, adding its replicates causes a change of the first PC, otherwise no change of direction occurs. Hence, the osPCA AS is defined by the cosine similarity:

$$s_t = 1 - \left| \frac{\langle \tilde{u}_t, u \rangle}{\|\tilde{u}_t\| \cdot \|u\|} \right|,$$

where u is the direction obtained using the dataset without the point x_{new} while $\tilde{u}_t$ represents the direction obtained using the dataset with the point x_{new} replicated $\tilde{n}$ times. Aiming at alleviating the PCA computation complexity, the first PC $\tilde{u}_t$ is approximated with a method called Power Method [9]; Power Method is an iterative procedure that allows approximating the first PC in m steps.

The main drawbacks of osPCA are:

- It does not guarantee a fast convergence, even if we use prior principal directions as its initial solutions;
- It requires the user to keep the entire covariance matrix, which prohibits the problems with high-dimensional data or with limited memory resources.

OnlinePCA [8] aims at overcoming the limitations of osPCA approximating the PCA computation to a Least Square problem. More concretely, after some passage, the PCA problem

$$\min_{U \in \mathbb{R}^{p \times k}, \|U\|=I} \sum_{i=1}^{n} \|(x_i - \mu) - UU^T(x_i - \mu)\|^2$$

is approximated by

$$\min_{\tilde{U} \in \mathbb{R}^{p \times k}, U^T U = I} J_{ls}(\tilde{U}) \approx \sum_{i=1}^{n} \left\|\bar{x}_i - \tilde{U} y_i\right\|^2 + \tilde{n}\left\|\bar{x}_t - \tilde{U} y_t\right\|^2,$$

where U is computed (once) in advance. The first PC is obtained by

$$\tilde{u} = \frac{\beta(\sum_{i=1}^{n} y_i \bar{x}_i) + y_t \bar{x}_t}{\beta\left(\sum_{i=1}^{n} {y_i}^2\right) + y_t^2}$$

and the onlinePCA AS is computed as in osPCA.

2.2. Angle Based Outlier Detection (ABOD)

Angle Based Outlier Detection (ABOD) [10] is an angle-based method, where angles are computed between an observation in exam and other samples available in the dataset. The intuition behind ABOD, is that inliers will produce angles with high variance since they are inside a "cluster", while outliers will have associated angles with low variances since they are outside a "cluster"; this concept is illustrated in Figure 1.

More specifically, the SA s_{ABOD} is defined by the variance over the angles between the difference vectors of x to all pairs of points in X weighted by the distance of the points:

$$s_{ABOD}(x) = Var_{y,z \in D}\left(\frac{\langle \overrightarrow{xy}, \overrightarrow{xz} \rangle}{\|\overrightarrow{xy}\|^2 \cdot \|\overrightarrow{xz}\|^2}\right) \geq 0, \qquad x, y, z \in \mathbb{R}^p$$

The distance weighting is used to enhance the fact that outliers are commonly distant to normal attributes. In fact, a low s_{ABOD} means that x is an outlier. A great advantage of ABOD is being a parameter-free procedure, however this comes at the cost of an high complexity: $O(n^3p)$. In order to alleviate this issue, a two-step approximation is employed, called *Lower Bound-ABOD (LB-ABOD)* based on a 6-steps procedure:

1. For each point $x \in X$, derive the k nearest neighbors [11].
2. Compute $s_{LB-ABOD}$ for each point $x \in X$.
3. Organize the database objects in a candidate list ordered ascendingly w.r.t. their assigned $s_{LB-ABOD}$.
4. Determine the exact s_{ABOD} for the first l objects in the candidate list, remove them from the candidate list and insert them into the current result list.
5. Remove and examine the next best candidate c from the candidate list and determine the exact s_{ABOD}. If the s_{ABOD} of c is smaller than the largest s_{ABOD} of an object x in the result list, remove x from the result list and insert c into the result list.
6. If the largest s_{ABOD} in the result list is smaller than the smallest $s_{LB-ABOD}$ in the candidate list, terminate. Else, proceed with step 5.

The procedure is motivated by the following fact. Suppose an attribute $x \in \mathbb{R}^p$ is the first-rank outlier, then if an attribute is an outlier also its $s_{LB-ABOD}$ is high. This method has the advantage of not being purely distance-based, which makes it more suitable for high-dimensionality w.r.t. distance-based methods. However, it may not capture complex structures (it requires a "cluster") and its complexity is $O(p \cdot n^2)$.

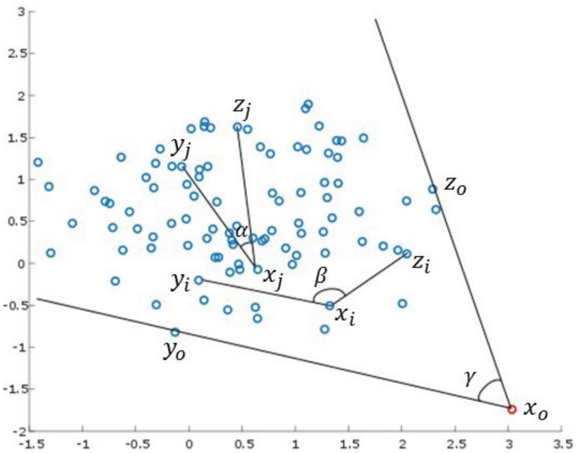

Figure 1. The ABOD procedure. Inliers (such as x_i and x_j) form high variance angles with other samples in the dataset, while outliers (like x_o) form low variance angles with other samples in the dataset.

Local Outlier Factor (LOF)

Local Outlier Factor (LOF) [12] is a density-based technique, which considers local neighborhoods to compute the anomaly score. An outlier is characterized by having low-density neighborhoods, that is, there are very few attributes in its neighborhood. This method, like ABOD, assumes that normal instances are organized in clusters (possibly more than one and with different density).

Now, let $d_k(x)$ be the distance from $x \in \mathbb{R}^p$ of the k-nearest neighbor and let $r_k(y,x) = max\{d_k(x), d(y,x)\}$ be the *reachability distance*, which is introduced to reduce statistical fluctuation in the computation of s_{LOF}. As said, an attribute x is characterized in LOF by its local density, which is called *Local Reachability Density* lrd_k and defined by the inverse of the average reachability distance based on the k-nearest neighbors $N_{k(x)}$:

$$lrd_k(x) = \left(\frac{\sum_{y \in N_{k(x)}} r_k(x,y)}{k}\right)^{-1}.$$

The anomaly score of LOF for a point x is defined by:

$$s_{LOF}(x,k) = \frac{\sum_{y \in N_{k(x)}} \frac{lrd_k(y)}{lrd_k(x)}}{k} \in [0,1]$$

and represents the degree of 'local' anomaly. When the local density around x is much lower than those of the other neighbor points $y \in N_{k(x)}$, $s_{LOF}(x,k)$ will tend to 0, meaning that it is an outlier; otherwise if the local density around x is very similar to those of the other neighbor points $y \in N_{k(x)}$, $s_{LOF}(x,k)$ will tend to 1, meaning that it is "inside" a cluster.

Although the only parameter to be tuned in LOF is k, its relation with LOF is very complex, which make difficult to find a heuristic to estimate the optimal k. In order to overcome this issue, it is suggested to select two bounds kLB (lower bound) and kUB (upper bound) and ranking all objects with respect to the maximum LOF value within the specified range and take the maximum. For further details on lower and upper bound of LOF we refer the interested reader to [12].

LOF is a computationally expensive with the complexity $O(kn^2p)$. Another drawback of LOF is that its performances are poor with low-dimensional structures or more complex structures than simply distributed clusters.

2.3. Comparison

The compared approaches are summarized in Table 1.

Table 1. Complexity comparison of the considered Anomaly Detection approaches

Method	Tuning Parameter	Training Complexity	Evaluation Complexity
osPCA	m	$O(mnp)$	$O(mnp)$
onlinePCA	m	$O(mnp)$	$O(p)$
ABOD	-	$O(n^3p)$	$O(n^3p)$
LB-ABOD	k	$O(n^2p + k^2p)$	$O(n^2p + k^2p)$
LOF	k	$O(kn^2p)$	$O(kn^2p)$

3. Semiconductor Manufacturing Case Study: Etching

The compared anomaly detection strategies have been tested on a real industrial dataset related to a Semiconductor Manufacturing Etching process [13]. Semiconductor Manufacturing fabrication is based on wafers. A wafer is a thin (125 - 300mm diameter and 275 - 775 μm) slice of semiconductor material - usually silicon crystal - that serves as the substrate for microelectronic devices. Production is often organized on *lots*, set of wafers (usually 25) that are typically moved across the fab and processed jointly/in the same equipment.

The data available consists of 2194 wafers belonging to a set of 313 (not complete) lots, for which Optical Emission Spectrometry (OES) are available. The process in question is plasma etching, for which OES data represents a non-costly and informative source of information from a chemical point of view. The dataset also includes information on another important quantity, the Etching Rate that is the ratio between the depth of the created trench and the time taken to perform the 'excavation'.

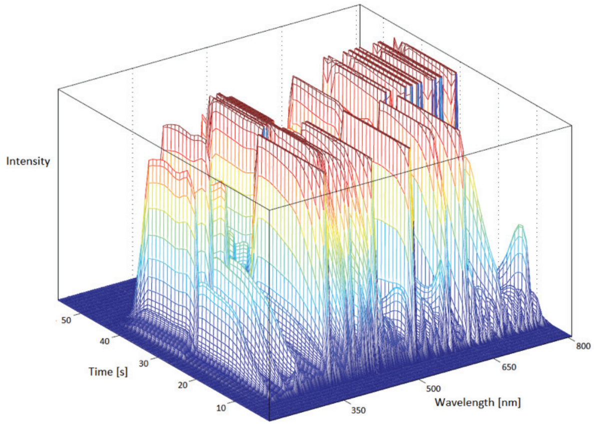

Figure 2. Typical OES data.

The Etch Rate is used in this work as a quality indicator for the produced wafers: outliers or changes seen in the Etch Rate can be considered as anomalies or changes in the production. Unfortunately, the Etch Rate is costly to compute in production and in some production settings is not available for most produced wafers; this scenario underlines the importance of an AMS that is able to infer anomalous situations from different data sources, like OES data. The available dataset consists of 2170 inlier wafers and 24 outlier wafers, labeled as such by inspecting the Etch Rate. From such data, features are extracted to be employed in the aforementioned Anomaly Detection procedures; such features are mean and variance over time for each of the 2048 monitored wavelengths in the OES data for a total of 4096 features considered. A data reduction procedure based on eliminating redundant features that have correlation above 0.99 is then performed: the final number of features considered is p= 997.

3.1. Experimental Results

The proposed methodologies have been compared through a Monte Carlo Cross-Validation (MCCV) procedure (also known as Repeated Random Sub-sampling Validation) [14]; the $N = 2194$ have been divided into:

- *training data* (50%), a set of observations on which the thresholds of the τ ASs are tuned; such set of data is also used to find the optimal values of the
- *validation data* (the remaining 50% of data), used for evaluating the performances of the compared methodologies.

The split between training and validation is done at random, preserving the ratio between outlier and inlier observations in each set (stratified cross-validation). The aforementioned procedure is repeated $K = 100$ times, by randomly choosing how data are assigned to the training or validation set: results are reported in the following as average over the K MCCV simulations.

The results are reported in terms of two indicators:

- *precision*, the ratio between true outliers and total outliers (both true and false);
- *recall*, the ratio between true outliers and true outliers plus false inliers.

The experimental results are reported in Table 2; in the LOF family of approaches, only LOF have been considered, given its higher expected performances. From Table 2 it can be appreciated how all the methods guarantee above 76% precision and above 89% recall. Moreover, it can be appreciated how the best performances were achieved by the LOF, in terms of precision, and by the ABOD, in terms of recall. In Table 2 is also reported the number of times each method has achieved the highest precision and the number of times it has achieved the highest recall (over the $K = 100$ MCCV); in case of a tie between more than one method, such simulation is counted for each methods in these indicators,

Table 2- Performances of the considered Anomaly Detection approaches with the Semiconductor Manufacturing dataset

Method	Average Precision over K = 100 MCCV simulations	# MCCV simulation with highest precision	Average Recall over K = 100 MCCV simulations	# MCCV simulation with highest recall
osPCA	76.43%	6	89.17%	12
onlinePCA	76.47%	6	89.08%	12
ABOD	76.32%	4	$\mathbf{91.33}$%	**53**
LOF	$\mathbf{79.08}$%	**87**	90.42%	31

4. Conclusions

In this paper, several approaches for Anomaly/Outlier detection have been compared on a real industrial case study; the case study in exam was related to Etching, one of the main step in the Semiconductor Manufacturing fabrication. It has been shown how monitoring the OES data provided satisfying results in detecting outliers; on the dataset at hand ABOD outperformed the other methods in terms of recall, while LOF was the anomaly detection approach with the highest precision. Future works will regard more sophisticated approaches to extract features from OES data; moreover, new policies for setting up the threshold τ on the anomaly score will be investigated.

References

[1] B. Haddad, S. Yang, L. Karam, J. Ye, N. Patel, M. Braun, Multifeature, Sparse-Based Approach for Defects Detection and Classification in Semiconductor Units, vol (xx), available online http://ieeexplore.ieee.org/document/7557053/

[2] H. Wang, F. Tan, B. Sheng, J. Bian, T. Pan. Run-to-run Control for Semiconductor Manufacturing Processes using Extended State Observer. In IEEE Chinese Control and Decision Conference (2016), pp. 854-857.

[3] Y. Zheng, D. Ling, Y.-W. Wang, S.-S. Jang, B. Tao, Model Quality Evaluation in Semiconductor Manufacturing Process with EWMA Run-to-Run Control, (2017), vol. 30(1), pp. 8-16.

[4] G.A. Susto, A. Beghi, Dealing with Time-series Data in Predictive Maintenance Problems, IEEE Emerging Technologies and Factory Automation Conference (2016).

[5] A. Beghi, L. Cecchinato, C. Corazzol, M. Rampazzo, F. Simmini, G.A. Susto, A One-Class SVM Based Tool for Machine Learning Novelty Detection in HVAC Chiller Systems, 19th World Congress of the International Federation of Automatic Control (2014), pp. 1953-1958.

[6] H. Rostami, J. Blue, C. Yugma. Equipment Condition Diagnosis and Fault Fingerprint Extraction in Semiconductor Manufacturing. Fifteenth IEEE International Conference on Machine Learning Applications (ICMLA) (2016).

[7] Y.-J. Lee, Y.-R. Yeh, Y.-C. Wang, Anomaly Detection via Online Oversampling Principal Component Analysis, IEEE Transactions on Knowledge and Data Engineering (2013), vol. 25(7), pp. 1460-1470.

[8] I. Joliffe, Principal Component Analysis, John Wiley & Sons, Ltd (2002).

[19] M. Journée, Y. Nesterov, P. Richtárik, R. Sepulchre, Generalized Power Method for Sparse Principal Component Analysis, Journal of Machine Learning Research (2010), pp. 517-533.

[10] H.-P. Kriegel, A. Zimek, Angle-based Outlier Detection in High-Dimensional Data, Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008), pp. 444-452.

[11] G.A. Susto, A. Schirru, S. Pampuri, S. McLoone, A. Beghi, Machine Learning for Predictive Maintenance: a Multiple Classifiers Approach, IEEE Transactions on Industrial Informatics, vol. 11 (2015), pp. 812–820.

[12] M. Breunig, H. Kriegel, R. Ng, J. Sander, LOF: Identifying Density-based Local Outliers, ACM Sigmoid Record (2000), vol. 2, pp. 93-104.

[13] S.J. Hong, W.Y. Lm, T. Cheong, G. May, Fault Detection and Classification in Plasma Etch Equipment for Semiconductor Manufacturing e-Diagnostics, IEEE Transactions on Semiconductor Manufacturing, vol. 25 (2012), pp. 83-93.

[14] J. Friedman, T. Hastie, R. Tibshirani. The elements of statistical learning. Springer, 2009.