Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-5 (2016) 132–137

Feature Selection for Anomaly Detection Using Optical Emission Spectroscopy

Luca Puggini * **Seán McLoone** **

* *National University of Ireland Maynooth, Maynooth, Ireland (e-mail: lpuggini@eeng.nuim.ie)*
** *Queen's University Belfast, Belfast, England (e-mail: s.mcloone@qub.ac.uk)*

Abstract: To maintain the pace of development set by Moore's law, production processes in semiconductor manufacturing are becoming more and more complex. The development of efficient and interpretable anomaly detection systems is fundamental to keeping production costs low. As the dimension of process monitoring data can become extremely high anomaly detection systems are impacted by the curse of dimensionality, hence dimensionality reduction plays an important role. Classical dimensionality reduction approaches, such as Principal Component Analysis, generally involve transformations that seek to maximize the explained variance. In datasets with several clusters of correlated variables the contributions of isolated variables to explained variance may be insignificant, with the result that they may not be included in the reduced data representation. It is then not possible to detect an anomaly if it is only reflected in such isolated variables. In this paper we present a new dimensionality reduction technique that takes account of such isolated variables and demonstrate how it can be used to build an interpretable and robust anomaly detection system for Optical Emission Spectroscopy data.

© 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Semiconductors, Fault Detection, Dimensionality Reduction, OC-SVM, OES Spectrum

1. INTRODUCTION

Semiconductor manufacturing is one of the most rapidly evolving industries. To remain competitive producers must continually deliver new devices that are smaller, faster and/or more energy efficient that previous generations, while at the same time keeping production costs low. In this context the ability to detect faults during the production process reduces the number of incorrectly processed wafers and directly translates into improved overall process yield and throughput (He and Wang (2007)).

As a result, fault or anomaly detection is an active area of research within the semiconductor manufacturing environment. Some recent examples are Puggini et al. (2015) and Mahadevan and Shah (2009) where anomaly detection in OES time series is performed with unsupervised random forest and one class support vector machines (OC-SVM) or Ren and Lv (2014), He and Wang (2007) and Verdier and Ferreira (2011) where clustering is used to separate normal and anomaly samples.

Data driven anomaly detection systems can roughly be divided into three subgroups according to the information available about the data during the training phase:

- *Supervised anomaly detection* where samples from normal and abnormal behaving wafers are available to train classifiers such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and k-nearest neighbours to distinguish between normal and anomalous samples (Chandola et al. (2009)).
- *Semi-supervised anomaly detection* where only data for normal samples is available. Systems can then be trained to assign an anomaly score to new samples according to how distant they are from the normal behaving ones. Several algorithms have been developed with this aim, including Multivariate Control Charts (Lowry and Montgomery (1995)), one-class SVMs (Schölkopf et al. (2001)) and Unsupervised Random Forests (Shi and Horvath (2006)).
- *Unsupervised anomaly detection* where no information is available about the data (i.e. the data is unlabeled) but assumptions are made regarding the frequency and distinctiveness of the anomalies within the overall dataset. This structure is then revealed and potential anomalies identified though the application of unsupervised clustering techniques such as DBSCAN (Ester et al. (1996)) and Max Separation clustering (Flynn and McLoone (2011)).

Optical emission spectroscopy (OES) is increasingly being used by semiconductor manufacturers for plasma etch process monitoring due to its ability to track variations in the chemical composition of a plasma over time. The OES data is composed of measurements of the light emitted from the plasma as a function of wavelength and time. Figure 1 shows a sample spectrum from the plasma etch process case study which will be introduced in Section 3. OES has been shown to be an effective wafer processing monitoring signal (Chen et al. (1996), Puggini et al. (2014)) and has been employed for applications such as anomaly detection (Puggini et al. (2015), Yue et al.

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.102

(2000)) and etch rate prediction (Puggini and McLoone (2015), Zeng and Spanos (2009)). OES data is generally characterized by high dimension, (Prakash et al. (2012)) which poses a problem for anomaly detection algorithms. Most anomaly detection algorithms are based on a distance measure and it is known that distance measures become meaningless in high dimensional spaces due to the so-called curse of dimensionality (Kriegel et al. (2008)).

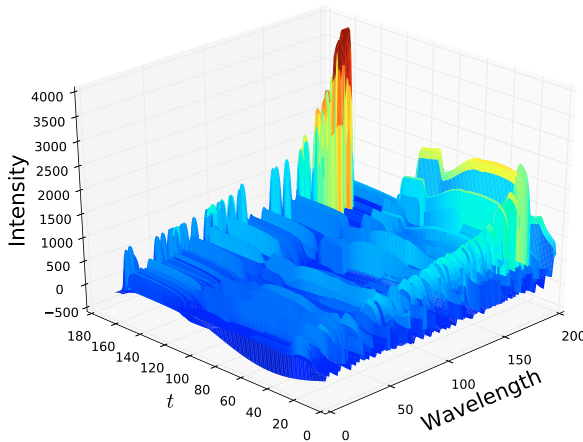

Fig. 1. A typical of OES spectrum from the case study presented in Section 3

In this paper the focus is on developing an appropriate data representation and dimensionality reduction technique for anomaly detection using OES in semiconductor manufacturing. In particular, a Forward Selection Independent Variables (FSIV) algorithm is proposed as an enhancement to Forward Selection Component Analysis (FSCA) (Prakash et al. (2012)) that yields better features for anomaly detection than Principal Component analysis (PCA) (Jolliffe (2002)) or FSCA when the anomaly occurs in an isolated variable in high dimensional correlated datasets. The efficacy of FSIV is demonstrated using both simulated and industrial case studies. In the industrial case study, a semi-supervised anomaly detection system is developed using a one-class SVM as the classification engine.

The remainder of the paper is organised as follows. Section 2 introduces the FSIV algorithm and demonstrates it performance with respect to PCA and FSCA for a simulated example. Similar results ar then presented in Section 3 for an industrial plasma etch case study. The anomaly detection classifier is developed in Section 4 and the results of its application to the industrial case study presented in Section 5. Finally, conclusions are provided in Section 6.

2. DIMENSIONALITY REDUCTION IN ANOMALY DETECTION

Dimensionality reduction techniques such as PCA and FSCA seek to obtain lower dimensional approximations of datasets from which it is possible to reconstruct the majority of the information in the original high dimensional datasets, usually defined in terms the percentage of explained variance. While they are generally very useful for generating compact representations of highly correlated datasets, the reduced representations are not guaranteed to retain sufficient information to detect isolated anomalies. In particular, in datasets with several large clusters of correlated variables, the contributions of isolated uncorrelated variables to explained variance may be insignificant, with the result that such variables may not be included in the reduced data representation. It is then not possible to detect an anomaly if it is only reflected in such isolated variables.

Mitra et al. (2002) and Flynn and McLoone (2011) have developed algorithms that perform unsupervised features selection while at the same time attempting to retain isolated variables in the data. In these algorithms the variables are recursively clustered. In the former for each variable the set of its k-nearest variables is computed according to a similarity function. The variable which is closest to its kth neighbour is retained while its k neighbours are discarded. The process ends when all the k-neighbours of all the variables are closer than a certain threshold to their centroid. In the latter centroids for new clusters are chosen based on how different they are from the data in existing clusters, and individual clusters are formed on the basis of exceeding a similarity threshold. Then when clustering is complete the reduced dataset representation is defined as the centroids of the clusters.

2.1 FSIV Algorithm

Both Mitra et al. (2002) and Flynn and McLoone (2011) select features based on a function $s(x, y)$ that measures the similarity between two variables. In general, instead of discarding variables that are similar to those already selected, it is more interesting to know which variables are not adequately represented by the selected variables. With this in mind Forward Selection Independent Variable (FSIV) analysis is proposed as a tool for efficient unsupervised features selection in anomaly detection.

Here the steps required to select K variables with FSCA (Prakash et al. (2012)) are recalled:

1 Start with the full data $X = (x_1, \ldots, x_p)$ and K the number of variables to select. Initialize $Z_0 = \emptyset$ and $k = 0$.
2 Scale the data to zero mean.
3 Define Z^v_{k+1} as the matrix Z_k with the addition of the variable x_v i.e. $Z^v_{k+1} = (Z_k, x_v)$
4 Define Z_{k+1} as:
$$argmin_v \parallel X - Z^v_{k+1}(Z^T_{k+1^v} Z^v_{k+1})^{-1} {Z^v_{k+1}}^T X \parallel_2 \quad (1)$$
5 Update $k = k + 1$
6 If $k < K$ return to step 3. Otherwise output Z_K, the set of selected variables.

The FSIV algorithm begins by selecting its first k variables $(z_1, \ldots, z_k)$ using the FSCA algorithm. This step is required to ensure the presence of the variables that represent the largest variation in the data. Then, additional variables are added in order to model signficant isolated variations that are not captured by the first k variables. The process ends when K variables are selected or when the error $\epsilon_{\hat{j}}$ defined according to equations 4 and 5 is smaller than a given threshold. The FSIV algorithm is thus defined as follows:

Download English Version:

https://daneshyari.com/en/article/710445

Download Persian Version:

https://daneshyari.com/article/710445

Daneshyari.com