[Sparse Modeling for Automatic Extraction of Variables to build accurate Virtual Metrology Models]

[Sumika ARIMA, Yuri ISHIZAKI, Huizhen BU]
[arima, s1620451, s1620528@sk.tsukuba.ac.jp]
University of Tsukuba
1-1-1 Tennodai, Tsukuba city, Ibaraki pref., Japan
Phone: +81 -298-535-578 Fax: +81-298-535-558

ABSTRACT

This paper introduced applications of sparse modeling (SPM) to automatically extract significant variables of equipment condition to estimate the product quality. Particularly, Lasso technique of SPM was effectively used before building a virtual metrology (VM) model of a kernel Support Vector Machine (kSVM). The scale of the variable's set was compressed by almost 60%, and 3 sensors can be reduced without decrease of the VM model accuracy. Lasso-kSVM can become a robust VM tool.

I. INTRODUCTION

In 2015, we have presented applications both of data mining and machine learning methods for AEC/APC; automatic fault detection and classification (FDC) of a discrete process, and highly accurate virtual metrology (VM) of a low-volume product-mix production system [1]. As the first, VM of a plasma-CVD process was examined by using single control charts, some multivariate statistical analyses, and a support vector machine (SVM) which is known as an accurate machine learning method [2]. Much adaptive accuracy of kernel-SVM could be evaluated by numerical experiments of discrimination of multi-class quality by using actual fab data. However, 13 different variable sets are comparably evaluated to get the best accuracy in that case. The number of the variables (M=90) are larger than the number of samples (e.g. n=50), and thus the some of the variables (36) are selected based on K.K.D.

Here, the sparse modeling is a rapidly progressing in recent [3][4]. It is one important research area of the compressed sensing, and it has very wide application fields such as a medical data processing (rapid image sensing of MRI or CT), the earth science (data-driven modeling and forecasting), and so on. Note that the deep learning method also can extract meaningful variables but it requires a big data to analyze, and so it cannot be used in this case. This paper focuses on the automatic variable extraction which can be used even when M >n.

II. DISCUSSION

A) METHOD – Sparse Modeling and Lasso

Basic Least-square method is used to estimate coefficients $\mathbf{C} \in \mathbb{R}^M$ to minimize the estimation error (Eq.2) in the linear regression (Eq.1) from the data set $\{(\boldsymbol{x}_i, y_i) \in \mathbb{R}^M \times \mathbb{R};\ i = 1,2,\cdots n\}$. Here, if some variables in $\boldsymbol{x}$ may not contribute to estimate Y, some of values in "C" should be zero to reduce the variance of the estimation result. That responds to "parse" case. Lasso (Least Absolute Shrinkage and selection operator) proposed by Tibshirani [5] is the estimation method of the sparse coefficient vector to reduce the variance of the estimation result (Eq.3).

$$Y = \boldsymbol{C}^T\boldsymbol{X}, \quad Y = (y_1, y_2, \cdots y_n)^T, \boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1{}^T \\ \boldsymbol{x}_2{}^T \\ \vdots \\ \boldsymbol{x}_n{}^T \end{pmatrix} \quad \cdots (\textbf{Eq. 1})$$

$$\boldsymbol{Er} = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i{}^T\boldsymbol{c})^2 = \|\boldsymbol{y} - \boldsymbol{X}c\|_2^2 \cdots (\textbf{Eq. 2})$$

$$\min_{c} \|\boldsymbol{y} - \boldsymbol{X}c\|_2^2 \quad \boldsymbol{s.t.} \|c\|_1 \le \boldsymbol{\lambda} \quad (\boldsymbol{\lambda} > \boldsymbol{0}) \quad \cdots (\textbf{Eq. 3})$$

B) APPLICATION

CVD process in a real mass production factory is a target of numerical experiments. Quality is measured by nine points on a wafer after the CVD process. Quality was categorized by 2 dimensions, uniformity and design conformity as shown in Table.1. On the other hand, 18 sensors are set to 4 subunits of the equipment to measure equipment conditions in the process. For every sensor data, five kinds of basic statistics (max, min, range, average, standard deviation(SD)) are computed soon after the equipment process completion. The equipment subunit cooperatively influences the process conditions, and thus the best accuracy of the virtual metrology is observed when all of the variables used together (Fig.1). However, it spends both time and cost to select the significant variables to use for VM when there are only small data (several dozen samples) measured under the same recipes such as high-mix low-volume production. In case of Fig.1, we empirically selected variables of 2 kinds of statistics as the result (Table.2). Here, we try to "automatically extract the variables by applying the sparse modeling, and evaluate those accuracy as well. The significant variables for design conformity

(V(*D*)) and for uniformity (V(*U*)) are extracted (Table.3, Figs.2,3). VM model of kSVM is built by using conjoint form of variables (e.g. V(D)∪V(U)) as the first. Its accuracy is evaluated for 2D quality discrimination by using a leave-one-out cross validation (Table.4). A linear kernel is the best for the Lasso-kSVM.

C) SUMMARY

This paper discusses the automatic extraction of significant variables to compress its sensing and data set effectively. 2-dimensional quality of product wafers should be estimated by less variables for low-volume production, and the Lasso technique of the sparse modeling (SPM) is applied before building the VM model using the support vector machine. 90 possible variables are defined to describe the process condition of Plasma-CVD equipment. The Lasso automatically extracts a few dozens of variables, and the scale of the variable's set is compressed by almost 60%. Also 3 sensors (#3, 5, 6) can be reduced without decrease of the VM model accuracy. In addition, each dimension of quality is described by different set of the variables including the same variables (B), however, its conjoint form well predicts a product quality for VM.

REFERENCES

[1] S. ARIMA, Y.-F. Wang, Y. ISHIZAKI, "Applications of machine learning and data mining methods for advanced equipment control and process control," Proceedings of AEC/APC symposium Asia 2015, DA-O-34 (pp.1-6), 2015.

[2] T. Lee, C. O. Kim, "Statistical Comparison of Fault Detection Models for Semiconductor Manufacturing Processes," IEEE Trans. on Semi. Manufacturing, Vol. 28, Issue: 1, pp.80-91, IEEE, Feb. 2015.

[3] CREST (Sparse Modeling) (M. Okada, et. al) http://sparse-modeling.jp/program/X00.html (2017/07/10).

[4] T. TANAKA : Mathematics of Compressed Sencing," IEICE Fundamentals Review, vol.4, issue 1, pp.39-4, 電子情報通信学会, 2010.

[5] R. Tibshirani, "Regression shrinkage and selection via the lasso," J. R. Statist. Soc. B, vol.58, no.1, pp.267-288, 1996.

Table.1 Quality of multi dimension (3x3=9 classes)

		Uniformity by wafer[a]			
		a	*b*	*c*	*All (abc)*
Design-conformity[b]	*A*	46	18	6	70
	B	14	2	8	24
	C	2	0	4	6
	All (*ABC*)	62	20	18	100

[a] *a*, *b* and *c* are uniformity levels in better rank order.

[b] *A*, *B* and *C* are conformity levels in better rank order. *A*, *B* and *C* are conformity levels in better rank order.

Note: Numbers in table are probability of the sample data, colours means difference indication/actions to do).

Table.2 The set of equipment variables. (# of data) [1], only average and standard deviation combined manually.

Parameter set #	*Average*	*SD*	*Parameter set #*	*Average*	*SD*
1	All (18)	All (18)	8	Subunit# 3(7)	Subunit# 3(7)
2	All (18)	–	9	Subunit# 3(7)	–
3	–	All (18)	10	–	Subunit# 3(7)
4	Subunit# 1(6)	Subunit# 1(6)	11	Subunit# 4(4)	Subunit# 4(4)
5	Subunit# 1(6)	–	12	Subunit# 4(4)	–
6	–	Subunit# 1(6)	13	–	Subunit# 4(4)
7	Subunit# 2(1)	Subunit# 2(1)			

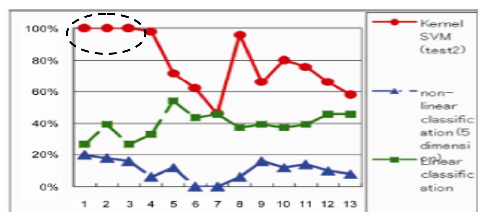

Fig. 1 VM accuracy –nine classes for 2D [1].
(axis X=1,..13 : No. the variable set in Table. 2)

Table.3 Variable extractions by Lasso technique - design conformity(D) / Uniformity (U) /Both (B)

Variables extracted		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
statistics	Min	U			U			U			U			U	U				
	Max	U						D				U	D	U	U	B	U	B	
	Range				U			D					U					U	U
	Average		B						U	U	B		B		U	D	U	U	U
	Std. dev.				U				U	B		U	B	U		U			

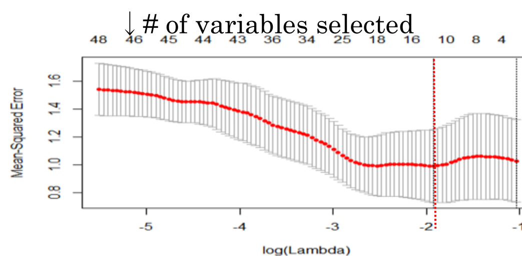

Fig. 2 Cross validation for Lasso parameter λ - i) ***D***
(λ - [min, lse] = [0.1462973, 0.3540583])

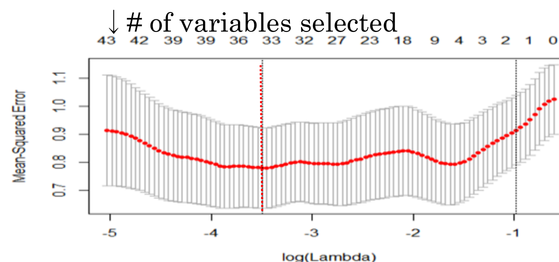

Fig. 3 Cross validation for Lasso parameter λ – ii) ***U***
(λ - [min, lse] = [0.03036876, 0.3744003])

Table.4 VM accuracy – Lasso-kSVM.

		Kernel	
Quality class of VM	**variables set [# of variables]**	**RBF**	**Linear**
Design conformity (A,B,C)	V(D) [11]	92	**88**
Uniformity (a,b,c)	V(U) [35]	84	**100**
2D quality (A,B,C) X (a,b,c)	V(D) U V(U) [39]	90	**100**
			accuracy [%]