Feature Selection and Boosting Techniques to Improve Fault Detection Accuracy in the Semiconductor Manufacturing Process

Kittisak Kerdprasop and Nittaya Kerdprasop, *Member, IAENG*

***Abstract*—Accurate and timely detection of any faults in the semiconductor manufacturing process is an important issue for process control engineers to ensure both productivity and reliability. Fault detection is a major step of process control aiming at constructing a decision tool to help detecting as quickly as possible any equipment or process faults in order to maintain high process yields in manufacturing. Traditional statistical based techniques such as univariate and multivariate analyses have long been employed as a tool for creating model to detect faults. Unfortunately, modern semiconductor industries have the ability to produce measurement data collected directly from sensors during the production process and such highly voluminous data are beyond the capability of traditional process control method to detect fault in a timely manner. We thus propose the techniques based on the data mining technology to automatically generate an accurate model to predict faults during the wafer fabrication process of the semiconductor industries. In such process control context, the measurement data contain over 500 signals or features. The feature selection technique is therefore a necessary tool to extract the most potential features. Besides the feature selection method, we also propose a boosting technique to handle the imbalance situation of fail versus pass test cases. The experimental results support our assumption that choosing the right features and boosting rare cases can considerably improve detection accuracy of fault products and processes.**

***Index Terms*—fault detection model, semiconductor manufacturing process, feature selection, rare case boosting**

I. Introduction

SEMICONDUCTOR manufacturing is a highly complex production process composed of hundreds of steps. The major processes in most semiconductor industries [1], [8] are in the following sequence: production of silicon wafers from pure silicon material, fabrication of integrated circuits onto the raw silicon wafers, assembly by putting the integrated circuit inside a package to form a ready-to-use product, and testing of the finished products. A constant advancement in the semiconductor industry is due mainly to persistent improvement of the wafer fabrication process. The fabrication process consists of a series of steps to cover special material layers over the wafer surface. Wafers reenter the same processing machines as each layer is successively covered. Some defects in this complicated process can make the final products fail the test. Fault detection and classification techniques [2], [3], [5-7], [11-15] applied to this critical manufacturing process can obviously improve product quality and reliability.

In recent years, many manufacturing tools are equipped with sensors to facilitate real-time monitoring of the production process. These tool-state and production-state sensor data provide an opportunity for efficient control and optimization. Unfortunately, such measurement data are so overwhelming that timely detection of any fault during the production process is difficult. In this paper, we study the problem of accurate detection of fault states in the wafer fabrication process. The dataset is donated by McCann *et al* [10] and publicly available for re-experimentation.

II. Related Work

Process control is crucially important to the semiconductor industries that operate the multistage manufacturing systems on the product scale of lesser 300 nanometers [12]. Modern technology in semiconductor manufacturing enables real time process control with the measurement data obtained from the equipment sensors and the final electrical test. With such high volume of data recorded during the entire production process, effective monitoring and optimal process control by investigating and analyzing these data are difficult work for process engineers. Traditional process control methodology like univariate and multivariate control charts is no longer an efficient method to control manufacturing systems with hundreds of processing stages. Instead automatic and advanced process control method is required.

Ison and colleagues [6], [7] proposed a decision tree classification model to detect fault of plasma etch equipment. The model was built from the five sensor signal data. Many researchers also studied the fault detection problem during the etch process. Goodlin *et al* [3] proposed to build a specific control chart for detecting specific type of faults. They collected tool-state data directly from the etcher. These data consist of 19 variables. The work of

Manuscript received December 8, 2010; revised January 20, 2011. This work was supported in part by grants from the National Research Council of Thailand, the Thailand Research Fund, and Suranaree University of Technology (via the Data Engineering and Knowledge Discovery (DEKD) Research Unit).

Kittisak Kerdprasop is the director of DEKD research unit and the associated professor of the school of computer engineering, Suranaree University of Technology, 111 University Avenue, Nakhon Ratchasima 30000, Thailand (phone: +66-(0)44-224349; fax: +66-(0)44-224602; e-mail: kerdpras@sut.ac.th, kittisakThailand@gmail.com).

Nittaya Kerdprasop is the co-founder and principal researcher of the DEKD research unit. She is an associate professor at the school of computer engineering, Suranaree University of Technology, Thailand (e-mail: nittaya@sut.ac.th).

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Spitzlsperger and colleagues [11] was also based on the statistical method. They adopted the multivariate control chart method to maintain changes in the mean and standard deviation coefficients by remodeling technique.

Recent interest in fault detection has been shifted toward the non-parametric approaches. He and Wang [5] proposed to use the k-nearest neighbor rule for fault detection. Verdier and Ferreira [14], [15] also applied the k-nearest neighbor method, but they proposed to use the adaptive Mahalanobis distance instead of the traditional Euclidean distance. Tafazzoli and Saif [13] proposed a combined support vector machine methodology for process fault diagnosis. Ge and Song [2] applied support vector data to the principal component analysis method to detect process abnormalities.

Most work on fault detection methods has studied the process control problem with a few features of tool-state and process-state measurement data. McCann and his team [9] proposed a rather different setting in which the measurement data from the wafer fabrication process contain as much as 590 features. With such abundant features or variables, feature selection techniques [4] are obviously necessary in order to improve both the prediction and the computational performances.

In this paper, we also analyze the wafer fabrication data [10] collected from 590 sensors with the last feature is a label stating pass or fail state. The observed data contain 1,463 pass cases with only 104 fail cases. In this work not only a feature selection method for extracting the post discriminative sensors is proposed, but also a boosting technique is devised to deal with highly imbalance between the pass and fail cases.

III. Fault Detection Technique

The SECOM dataset [10] contains 1567 examples taken from a wafer fabrication production line. Each example is a vector of 590 sensor measurements plus a label of pass/fail test. Among the 1567 examples, there are only 104 fail cases which are labeled as positive (encoded as 1), whereas much larger amount of examples pass the test and are labeled as negative (encoded as -1). The imbalance of pass and fail examples in addition to the large number of metrology data obtained from hundreds of sensors make this dataset a difficult one to accurately analyze. It is thus our main focus to devise a method based on data mining techniques to build an accurate model for fault detection. The framework of our study is presented in Fig.1.

Feature selection techniques in our study are ranging from simply removing features with a constant value and features containing numerous missing values (more than 55% of values are missing), to statistical based analysis such as chi-square and principal component analysis (PCA) and information theoretical based such as gain ratio. We also devise a cluster based technique call MeanDiff to analyze discrimination power of each feature. On the model building phase, we apply four methods to induce the fault-detection model namely decision tree, naïve Bayes, logistic regression, and k-nearest neighbor.

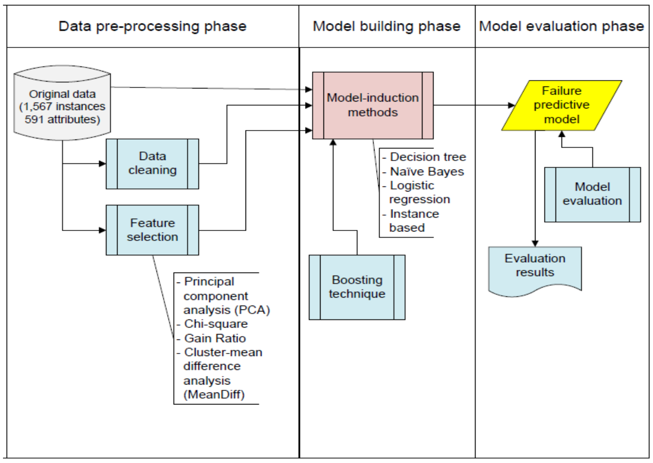

Fig. 1 Framework of proposed method and research study

The dataset is in a form of matrix; rows represent each observation or instance and columns represent features which are values recorded from each sensor. The steps in our proposed method for creating an accurate model to detect fault case from highly imbalance data with numerous features are as follows:

Data Cleaning Phase

(1) Investigate data observed from each sensor, i.e. data in each column. If the data appear to be a single value, then remove that feature.

(2) Count in each column the 'not available' or missing values. If data are missing more than 55%, then remove that feature.

Feature Selection Phase

(3) Apply two statistical based feature selection techniques: chi-square and principal component analysis (PCA), and save the result as two separate datasets.

(4) Apply an information theoretical based technique: gain ratio, and save the result in a separate dataset.

(5) Apply the following cluster-based feature selection technique, called *MeanDiff*:

(5.1) Clustering data into two clusters (fail cluster and pass cluster)

(5.2) Compare value differences in every feature of the fail cluster mean and the pass cluster mean

(5.3) Ranking features in descending order according to the magnitude of mean differences computed in step 5.2, and output the ranked features

Case Boosting Phase

(6) Separate data obtained from step 2 into two datasets: train data and test data. Each data set maintains the same proportion of pass and fail cases.

(7) Pumping the fail cases in the train data by duplicating the fail cases to be the same amount as the pass cases.

Model Building Phase

(8) Build a prediction model with decision tree, naïve Bayes, k-nearest neighbor, and logistic regression algorithms.

(9) For datasets from steps 3-5, evaluate model accuracy

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

with 10-fold cross validation technique. Dataset from step 7 is evaluated with the test set.

We assess the model performance based on the four metrics: true positive rate (TP rate or recall), precision, F-measure, and false positive rate (FP rate or false alarm). The computation methods of these metrics are given in Fig. 2 (TP = true positive, FP = false positive, FN = false negative, TN = true negative).

		Predicted class	
		Class=1 (fail)	Class= -1 (pass)
Actual class	Class= 1	TP	FN
	Class= -1	FP	TN

$$\text{TP rate, or Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{F-measure} = \frac{2TP}{2TP+FP+FN}$$

$$\text{FP rate} = \frac{FP}{FP+TN}$$

Fig.2 Fail/pass classification matrix and performance computation

IV. Experimental Results

A. *Feature Selection Technique Comparison*

We use the WEKA software [16] to perform a series of experiments. The first part of our study aims at selecting principal features that show the most discrimination power of differentiating fail cases from pass cases. In the cleaning step, we remove 137 features that contain a single value and lots of missing values. From the remaining 454 features, we select the best 168 features (to maintain around 95% of variances) by means of principal component analysis (PCA), Chi-square test, gain ratio computation, and our own MeanDiff method. The fault detection models are then derived from each feature selected data. We want the model that shows the highest values of TP rate, precision, and F-measure, but the lowest value in FP rate. The experimental results on the four model measurement metrics are shown in Figs. 3-6.

For this specific data domain, it can be noticed that feature selection can considerably improve the accuracy of fault detection models. The proposed MeanDiff method contributes the most to decision tree model, whereas the gain ratio method is the best feature selection method for the naive Bayes and logistic regression model building approaches. The k-nearest neighbor method (in which k was set to be one on our experiments because it yields the best result) needs a cleaned dataset without any other feature selection facility. If model comprehensibility is a major concern, the model built from a MeanDiff feature selected data with a decision tree approach is the most appropriate one. It is worth mentioning here that for such a large number of features like this application the neural network and support vector machine approaches consume so much memory that they cannot run to completion. Among the four model building methods, naïve Bayes model can detect fault cases at the success rate as high as 90%, but the false alarm (FP rate) is also as high as 80% as well. We compare the TP rate versus the FP rate of each model and provide the result in Fig. 7.

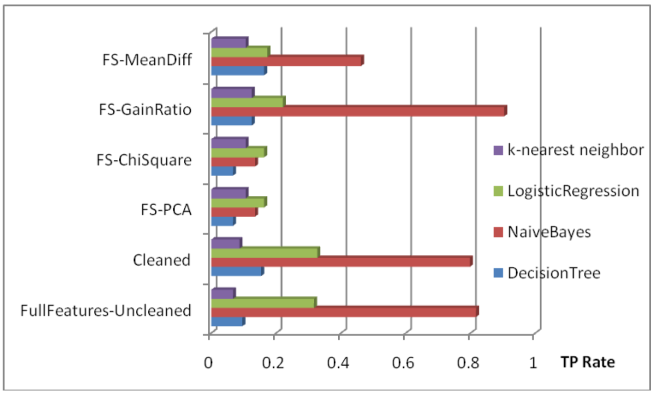

Fig. 3 TP rate of fault detection models on different feature selection methods

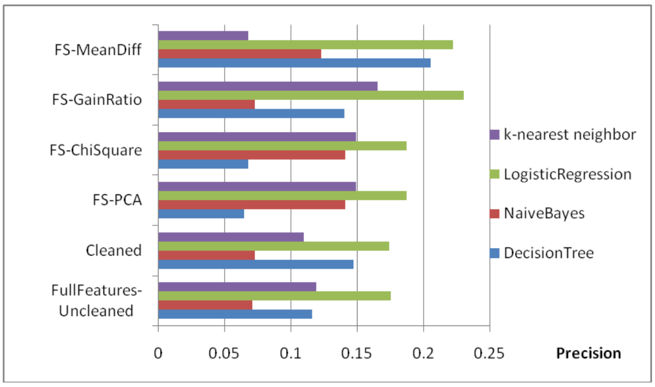

Fig. 4 Precision of fault detection models on different feature selection methods

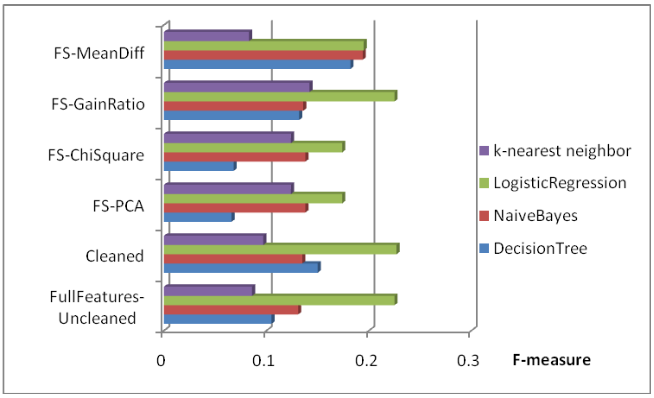

Fig. 5 F-measure of fault detection models on different feature selection methods

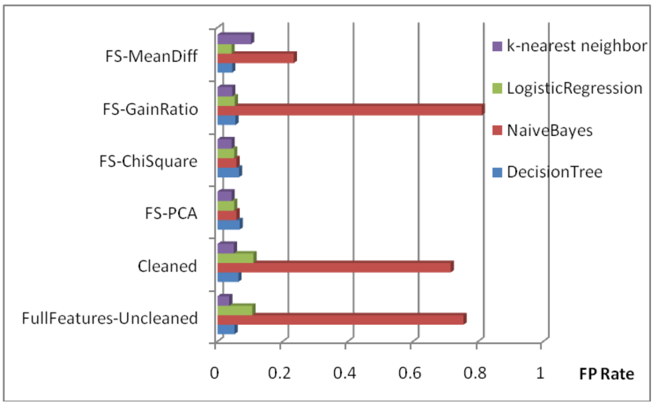

Fig. 6 FP rate of fault detection models on different feature selection methods

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

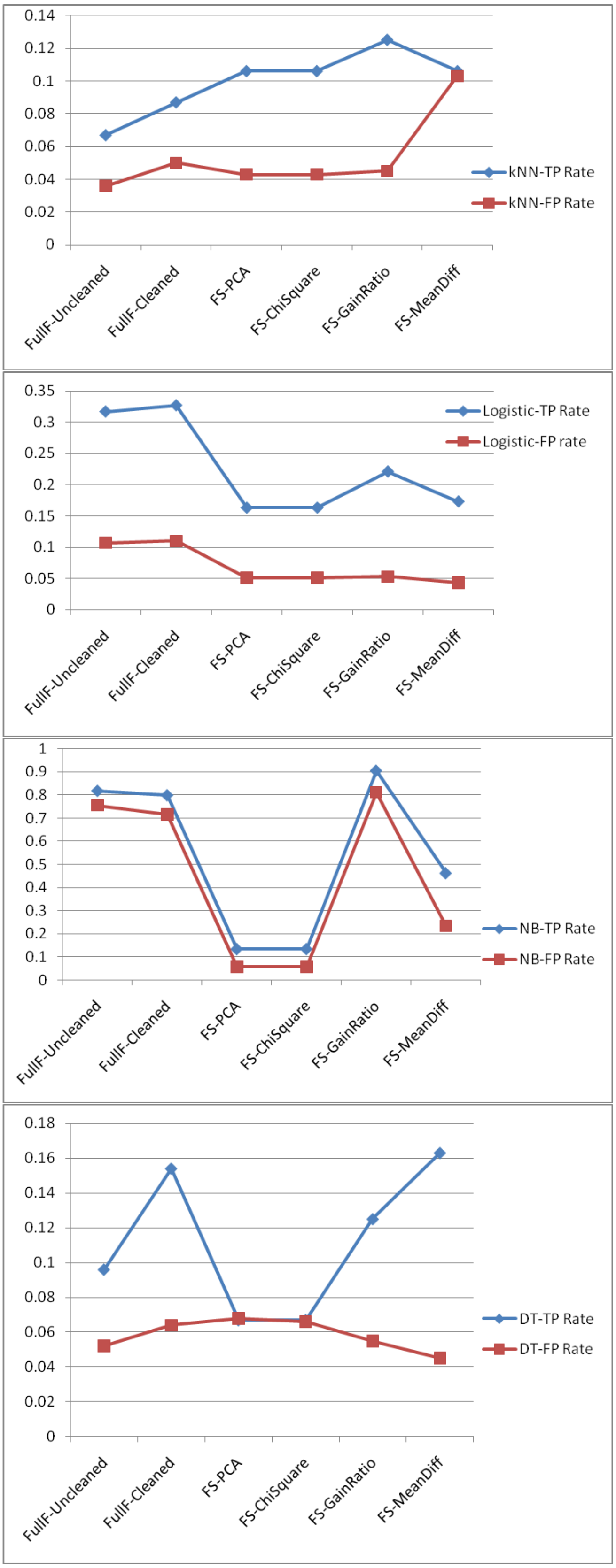

Fig. 7 TP rate versus FP rate comparison of each fault detection model

B. *Rare Case Boosting Results*

For the specific problem of fault detection, the number of fail test is very few (104 instances in the SECOM dataset) comparative to the number of pass test (1463 instances). It is therefore a difficult task to build automatically the accurate model that can detect such rare cases. We thus propose the idea of separating the SECOM dataset into a train set and a test set. The test set contains 468 instances in which 59 instances are fail test and 409 are pass test. The train set contains 45 instances of fail test and 1054 of pass test. We then duplicate the number of fail test in the training data to be 1096 instances. The fault detection models are built from this rare case boosting training dataset. The models are then evaluated their classification performances by the separated test dataset. The classification error matrices of models built from the four different learning methods are given in Fig. 8 and the performance criteria are summarized in Table 1. The boosted true positive rate, precision, F-measure, and the lower false positive rate of each model are also graphically provided in Fig. 9.

k-Nearest Neighbor — Predicted class

Actual class	Class=1 (fail)	Class= -1 (pass)
Class= 1	58	1
Class= -1	98	311

Logistic regression — Predicted class

Actual class	Class=1 (fail)	Class= -1 (pass)
Class= 1	59	0
Class= -1	137	272

Naïve Bayes — Predicted class

Actual class	Class=1 (fail)	Class= -1 (pass)
Class= 1	44	15
Class= -1	144	265

Decision Tree — Predicted class

Actual class	Class=1 (fail)	Class= -1 (pass)
Class= 1	59	0
Class= -1	66	343

Fig. 8 Classification error matrices of fault detection models

TABLE I
FAULT DETECTION MODEL ASSESSMENT

	k-Nearest Neighbor	Logistic Regression	Naïve Bayes	Decision Tree
TP rate	0.983	1.0	0.746	1.0
FP rate	0.24	0.335	0.352	0.161
Precision	0.372	0.301	0.234	0.472
F-measure	0.54	0.463	0.356	0.641

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

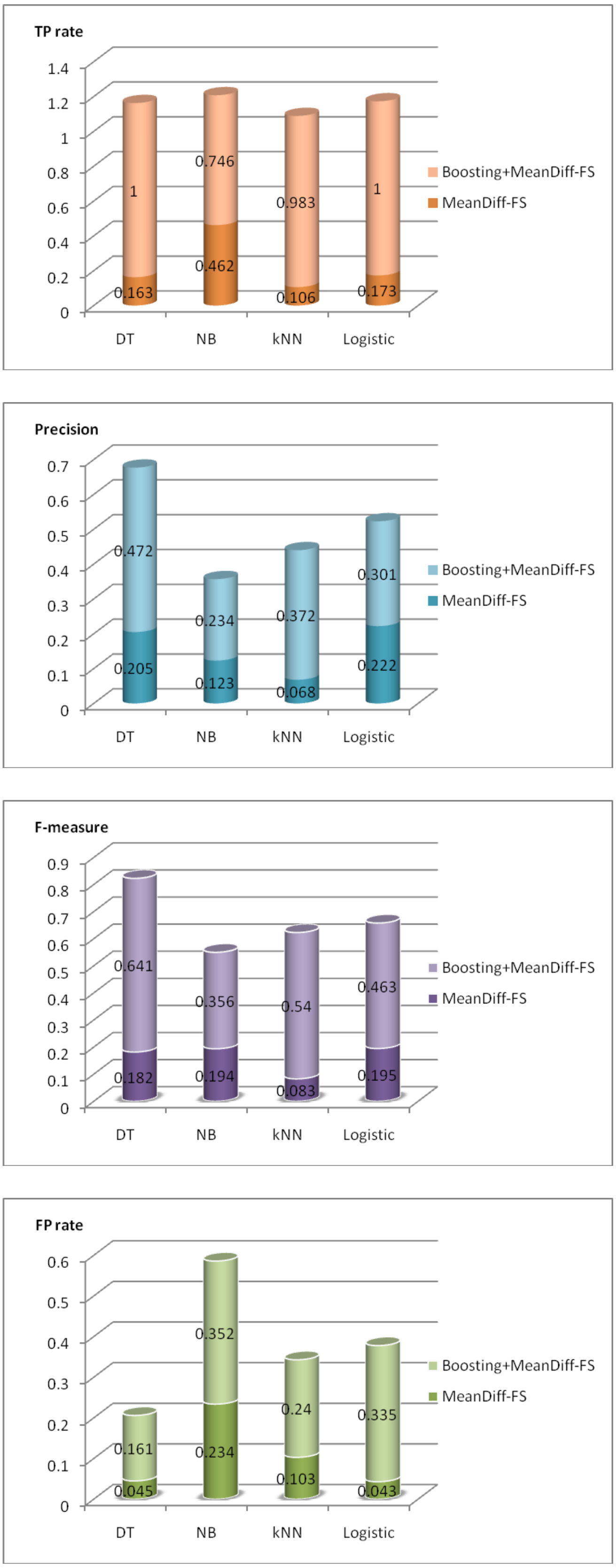

Fig. 9 High increases in TP rate, Precision, and F-measure but low increase in FP rate of fault detection models from applying the boosting technique

The high true positive but low false positive of the decision tree model make it a good candidate for automatic generation of the fault detection model to be used in the semiconductor manufacturing process. The fault detection model in a form of decision tree is given in Fig.10. The top level of the decision tree is on the left hand side in which the value from sensor number 511 is the first parameter to be considered. The normal state (encoded as -1) is expected if the value of sensor 511 is less than or equal 28.3784. The fault state is to be detected when the following sensor values are reported: S511 > 28.3784, S470 > 4.3751, S16 > 401.1307, S472 > 4.4751, S51 ≤ 646.9073, S4 > 905.1501, S188 > 11.54, S431 > 3.8926, S439 > 28.6219, S495 > 1.3638 S56 > 2875, S548 > 398.552, S178 ≤ 0.448, S29 ≤ 73.4556, S578 ≤ 16.4303, S474 ≤ 27.9511, and S39 ≤ 86.3506. Other prediction rules can be interpreted in the same manner.

S511 <= 28.3784: Predict -1
S511 > 28.3784
| S470 <= 4.3751: Predict -1
| S470 > 4.3751
| | S16 <= 423.3311
| | | S16 <= 401.1307: Predict -1
| | | S16 > 401.1307
| | | | S472 <= 4.4751: Predict -1
| | | | S472 > 4.4751
| | | | | S51 <= 646.9073
| | | | | | S4 <= 905.1501: Predict -1
| | | | | | S4 > 905.1501
| | | | | | | S188 <= 11.54: Predict -1
| | | | | | | S188 > 11.54
| | | | | | | | S431 <= 3.8926: Predict -1
| | | | | | | | S431 > 3.8926
| | | | | | | | | S439 <= 28.6219: Predict -1
| | | | | | | | | S439 > 28.6219
| | | | | | | | | | S495 <= 1.3638
| | | | | | | | | | | S56 <= 2875
| | | | | | | | | | | | S548 <= 398.552: Predict -1
| | | | | | | | | | | | S548 > 398.552
| | | | | | | | | | | | | S178 <= 0.448
| | | | | | | | | | | | | | S29 <= 73.4556
| | | | | | | | | | | | | | | S578 <= 16.4303
| | | | | | | | | | | | | | | | S474 <= 27.9511
| | | | | | | | | | | | | | | | | S39 <= 86.3506: **Predict 1**
| | | | | | | | | | | | | | | | | S39 > 86.3506: Predict -1
| | | | | | | | | | | | | | | | S474 > 27.9511: **Predict 1**
| | | | | | | | | | | | | | | S578 > 16.4303
| | | | | | | | | | | | | | | | S161 <= 614: Predict -1
| | | | | | | | | | | | | | | | S161 > 614: **Predict 1**
| | | | | | | | | | | | | | S29 > 73.4556
| | | | | | | | | | | | | | | S414 <= 25.0931: Predict -1
| | | | | | | | | | | | | | | S414 > 25.0931: **Predict 1**
| | | | | | | | | | | | | S178 > 0.448
| | | | | | | | | | | | | | S273 <= 19.8922: **Predict 1**
| | | | | | | | | | | | | | S273 > 19.8922: Predict -1
| | | | | | | | | | | S56 > 2875
| | | | | | | | | | | | S28 <= 7.373: Predict -1
| | | | | | | | | | | | S28 > 7.373: **Predict 1**
| | | | | | | | | | S495 > 1.3638: Predict -1
| | | | | S51 > 646.9073: Predict -1
| | S16 > 423.3311: Predict -1

Fig. 10 Decision tree model for fault-detection in the semiconductor process control

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. Conclusion

In semiconductor manufacturing process control and monitoring, hundreds of metrology data are available for process engineers to analyze for the purpose of maintaining efficient operations and getting optimum yield of high quality products. For such a large volume of measurement data, automatic fault detection technique is essential. We thus investigate the application of data mining techniques such as decision tree induction, naïve Bayes analysis, logistic regression, and k-nearest neighbor classification for creating an accurate model for fault case detection in the wafer fabrication process of semiconductor industries.

From a series of experimentation, we found that naïve Bayes model built from a subset of features selected by a gain ration criteria can detect the fault cases at the very high rate of 90%. But the false alarm rate, or false positive, is also as high as 80%. The decision tree method built from our MeanDiff feature selection method generates a more comprehensible form of fault detection model with false alarm rate at only 4.5%. But the precision and true positive rate, or recall, of the tree model are still low at 20.5% and 16%, respectively.

We thus devise a boosting technique to improve the precision of tree-based model for fault detection by pumping the number of rare cases, or fault test, to the equal number of majority cases, or pass test. The outcome is surprising that the true positive rate of the tree-based model can increase up to 100%, whereas the false alarm rate is still low at the 16%. We plan to investigate this boosting technique to other domains that show imbalance among data classes in our future research.

References

[1] E. Epistola. (2010, December 4). Semiconductor manufacturing. Available: http://www.siliconfareast.com/manufacturing.htm

[2] Z. Ge and Z. Song, "Semiconductor manufacturing process monitoring based on adaptive substatistical PCA," *IEEE Trans. Semiconductor Manufacturing*, vol.23, no.1, pp.99-108, Feb. 2010.

[3] B.E. Goodlin, D.S. Boning, H.H. Sawin, and B.M. Wise, "Simultaneous fault detection and classification for semiconductor manufacturing tools," *Journal of The Electrochemical Society*, vol.150, no.12, pp.G778-G784, 2003.

[4] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol.3, pp.1157-1182, 2003.

[5] Q.P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semiconductor Manufacturing*, vol.20, no.4, pp.345-354, Nov. 2007.

[6] A.M. Ison, W. Li, and C.J. Spanos, "Fault diagnosis of plasma etch equipment," in *Proc. IEEE Int. Symp. Semiconductor Manufacturing*, San Francisco, 6-8 October 1997, pp.B-49-B-52.

[7] A.M. Ison and C. Spanos, "Robust fault detection and fault classification of semiconductor manufacturing equipment," in *Proc. Int. Symp. Semiconductor Manufacturing*, Tokyo, 2-4 October 1996.

[8] G.S. May and C.J. Spanos, *Fundamentals of Semiconductor Manufacturing and Process Control*. John Wiley & Sons, 2006.

[9] M. McCann, Y. Li, L. Maguire, and A. Johnston, "Causality challenge: benchmarking relevant signal components for effective monitoring and process control," in *Proc. JMLR Workshop*, Canada, 12 December 2008, pp.277-288.

[10] SEmi COnductor Manufacturing. (2010, November 25). Available: http://www.causality.inf.ethz.ch/repository.php

[11] G. Spitzlsperger, C. Schmidt, G. Ernst, H. Strasser, and M. Speil, "Fault detection for a via etch process using adaptive multivariate methods," *IEEE Trans. Semiconductor Manufacturing,* vol.18, no.4, pp.528-533, Nov. 2005.

[12] S.J. Qin, G. Cherry, R. Good, J. Wang, and C.A. Harrison, "Semiconductor manufacturing process control and monitoring: a fab-wide framework," *Journal of Process Control,* vol.16, pp.179-191, 2006.

[13] E. Tafazzoli and M. Saif, "Application of combined support vector machines in process fault diagnosis," in *Proc. American Control Conf.*, St. Louis, USA, 10-12 June 2009, pp.3429-3433.

[14] G. Verdier and A. Ferreira, "Fault detection with an adaptive distance for the k-nearest neighbor rule," in *Proc. Int. Conf. Computer & Industrial Engineering*, Troyes, France, 6-9 July 2009, pp.1273-1278.

[15] G. Verdier and A. Ferreira. (2010). Adaptive Mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing. *IEEE Trans. Semiconductor Manufacturing*. Available: doi:10.1109/TSM.2010.2065531

[16] WEKA. (2010, October 23). Available: http://www.cs.waikato.ac.nz/ml/weka/

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)