International Journal of Engineering & Technology, 7 (2.33) (2018) 1229-1232

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

The Methodology of Sequences of Facilities using Production history information

Sung-Hyun Ryu[1], Sang-Hyun Choi[2*]

[1,2]Dept. of Management Information System, Chungbuk University,1 Chungdae-ro,Cheongju,Korea
**Corresponding author E-mail : chois@cbnu.ac.kr*

Abstract

Background/Objectives: In the 4th Industrial Revolution, innovation has been taking place in various fields in Korea. In particular, the concept of ‘Smart factory’ has come into being as a result of innovation in the manufacturing that is the foundation of the nation’s industry. However, small and meddle-sized companies don’t know how to set up an automated system and are reluctant to introduce it due to limitations of cost and so on.

Methods/Statistical analysis: Therefore, this paper proposes a sequences of facilities in the process that makes it easy for small and middle-sized enterprises (SMEs) to introduce 'Smart factory'. In this study, we choose the semiconductor process and use the processed production history data to derive the algorithm applying Conditional probability and Naive Bayesian. We utilized simulation data that have similar data type. We assumed the process of semiconductor and make the sample function using it.

Findings: We assume that the sequences of facilities in the processes may impact on the product qualities. So, we found out that the particular machines produced low-quality products through basic statistical graph. Moreover, it is possible to design the sequences of facilities that avoids producing low-quality products. As we apply the conditional probability and Naive Bayesian to the scheduling, performance is simple and fast to calculate by checking which equipment has passed and what quality of final product is. First, when one equipment is passed through the process, the equipment with the most defects is filtered, and the probability of a lot of defective products coming from the next equipment is sought. We can suggest the various scheduling methodologies proposed in the existing research. And compared to the existing research, less variables is needed to calculate further scheduling.

Improvements/Applications: For now, we suggest the algorithm. In the further study, we will apply the equation to the future research to construct the system. Also, we will systemize and apply the algorithm into it. It can show the entire process and we will distinguish which machine can produce high quality product.

Keywords: *Smart factory, Naive Bayes, Conditional probability, Production traceability, Algorithm development*

1. The Innovation of Manufacturing in the 4th Industrial Revolution

The Korean manufacturing industry is planning to innovate with various efforts for the government and enterprise in the4th Industrial Revolution. Traditional manufacturing methods have been limited by the decline in business performance, and the alternative was needed as it is undergoing a variety of environmental changes, including a decline in the manufacturing value and a shortage of workers. Accordingly, the smart factory's role in the Industry 4.0 has been highlighted as a way to overcome these crises. A smart factory is defined as one in which a system is built to automatically produce its own simulation machines by merging elements and IT infrastructures. Measuring or improving efficiency in various manufacturing processes is a major issue and has been studied in a variety of ways. [1]

So Byeong-Eob, Shin Sung-Sik(2017) said upgrading from a low to a high building level is effective for SMEs' introduction of smart factories. In addition, they emphasized that a 'gradual approach' should be attempted to build a successful smart factory. [2]

Until now, the 'gradual approach’ to smart factories for small and meddle-sized companies has been studied in a variety of ways. [3] Process scheduling is important because it directly has an influences on productivity and manpower utilization. Moreover, research is needed to improve the accuracy of future forecasting ability and reliable delivery dates. [4] Therefore, this study intends to focus on scheduling efficiency.

To overcome the limitations of SW scheduling, there has been a steady research on algorithms for automation and systematization.

Currently, most factories use production scheduling SW to design their production schedules, and the main related data are collected in a variety of ways, including ordering, master information, manufacturing instructions and performance information. However, the data that can be obtained from the production line are costly and time consuming to test operations. Also, there are several limitations not to be able to cope with the situation in real time during operation.[5]

In order to overcome those limitations of SW scheduling, there has been a constant research on algorithms for automation and systematization. An artificial neural network, which extracts data features from different factories and uses them to optimize their features, has been mainly used for scheduling efficiency studies.

Jo Yong-Cheol, Jo Hyeon-Chan, Kim Jong-Won, Jang Ryang, Jeon Hong-Tae(2008) said that , in the process of semiconductor production, the optimal scheduling of transport paths that occur during the Multi-spinner forming process will eventually reduce the waiting times and contribute to the production speed, if they use neural network.[6] As a result of the " generator case study, " Kim Wook, Park Young-Moon(1991) solved the problem of optimal route exploration with artificial neural networks to enhance the efficiency of the process.[7]

In addition, the study on manufacturing application using Naive Bayesian was conducted to improve the outflow rate and the ratio of the workload of human inspectors by selecting the main performance indicators in the outline inspection process and reflecting these characteristics to Naive Bayesian.[8]

In this study, the proposed scheduling algorithm is intended to be applied first to SMEs. Especially in factories producing semiconductors, it is very important to increase the yield, namely productivity.[9] To do this, we select a company that produces semiconductors in SMEs and design a virtual process to select necessary parameters. Finally, we propose an algorithm by applying data mining. In further study, we will prove the validity of the process scheduling algorithm that can improve the defect rate of the product and the efficiency of the work by applying the algorithm and verifying it to the real process.

2. Data Analysis with Simulation Data Set

2.1. Preprocessing of Simulation Data Set

The subject of this study was set up as a semiconductor line that went through complex processes, but the actual data could not be used for security reasons. So, we utilized and processed simulation data set that has similar data types with real data.

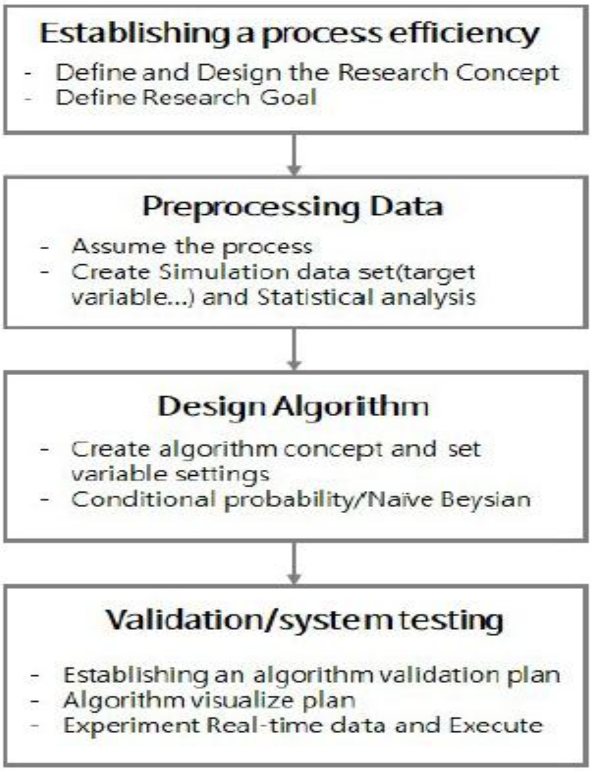

Figure 1:. The Process of Algorithm Development

We propose a process scheduling method as shown in Figure 1that can produce the best quality results using the variables generated from the SMEs' process. The data that can be generated by installing sensor in the process are extensive and diverse. In particular, it is used to derive algorithms to solve the problems of the schedule as described above by using conditional probability, Naive Bayesian and the production history data of the facilities. It could make the foundation for an optimized visualization system that will automatically guide the process paths that can produce high-quality results.

Our researchers would like to assume the following basic process scenarios in Figure 2. All final products go through a total of six processes and have different facilities/machine numbers for each process.

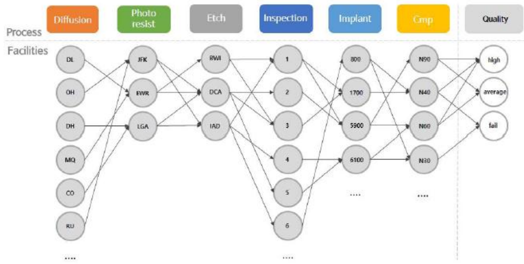

Figure 2:. The Process of Semiconductor

The assumed process consists of 2201 samples. The following data should be obtained for scheduling under the condition that all the final products throughout each process have the same process. The data set involve independent variable in table 1 including the equipment number assigned to each process and the target variable, 'Quality'.

Table 1:: Producing Process Variable

Producing Process Variable		
Name	**Variable type**	**Remarks**
Diffusion	DL, OH, DH, MQ, CO, RU, UA, US	Facility name corresponding to the Diffusion phase (8 facilities)
Photoresist	JFK, EWR, LGA	Facility name corresponding to the Photoresist phase (3 facilities)
Etch	BWI, DCA, IAD	Facility name corresponding to the Etch phase (3 facilities)
Inspection	1~7	Facility name corresponding to the Inspection phase (7 facilities)
Implant	5900,6100…	Facility name corresponding to the Implant phase (7 facilities)
Cmp	N94, N40…	Facility name corresponding to the Cmpphase (9 facilities)
Quality	high, average, fail	Product Quality

2.2. Data Basic Statistics Analysis

The basic statistical analysis of the studies is to examine the distribution of quality by facility for each process as shown in Figure 3. For the diffusion process, however, we checked that a facility with numbers "DH", "ACI", and "RU" is remarkable in the bad-quality products. Also, in the case of the Cmp process, it was confirmed that defects were found to be high in the equipment with the N60, N70, and N10 numbers. In later conditional probabilities and Naive Bayesian, these facilities will be used to predict quality by adding variables when determining badness.

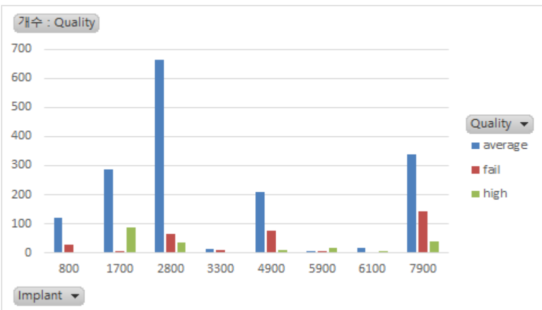

Figure 3:. The Basic Statistic Graph of the Number of Final Product in each process

3. The Suggestion of Algorithm about Semiconductor Process Scheduling

3.1. Conditional Probability

In this paper, we propose a new conditional probability and Naive Bayesian algorithm considering the conditions of equipment based on the previously designed semiconductor process model in Figure 4. Conditional probability generally means the probability that another event will occur under the condition that a particular event has occurred. Each product is identified as a result Q (high, average, fail) through a random f_k processes $(1 \leq f_k \leq K)$ and $M_{k,j}$facility(1 ≤ j ≤ the number of machine in each process). In the semiconductor process, there are constraints that the number of facilities operated by each process is different. To reflect this, the equipment number passing through the 'k' process is defined as $M_{k,j}$. The notation for algorithm development is defined as follows. Figure 4 shows the configuration of each process facility. The conditional probability is used to derive the probability of a defective product when passing through a specific process or facility, as shown in the following equation.

- $f_{k=1,2,3\ldots k\ldots K}$: Process number
- m_k : The number of machine that pass through k process
- $M_{k,j}$: machine that pass through k process
- Q : The production quality(Bad B/Good G/Average A)
- P(Q|$M_{k,j}$) : The probability when product quality that pass through $M_{k,j}$ process is Q (defective / good)

f_1	f_2	f_3	f_4	… k …	f_K
M_{11}	M_{21}	M_{31}	M_{41}	…	M_{K1}
M_{12}	M_{22}	M_{32}	M_{42}	…	M_{K2}
M_{13}	M_{23}	M_{33}	M_{43}	…	M_{K3}
M_{14}	M_{24}	M_{34}		$M_{k,j}$	M_{K4}
…	M_{25}	…		…	…
…		M_{3,m_3}		…	…
M_{1,m_1}					M_{K,m_K}

Figure 4:. The Semiconductor process

$$P(Q = B|M_{k,j}) = T\square e\ number\ of\ defective\ in\ M_{k,j}\ mac\square ine\ /\ Total\ number\ of\ Production\ t\square at\ pass\ t\square roug\square\ M_{k,j}\ mac\square ine \quad (1)$$

$$\min_{j=1,2,3\ldots j} P(Q = B|M_{k,j}) = \{M_{1,1}, M_{2,1}, M_{3,2} \ldots\}\ mac\square ine \quad (2)$$

Applying the above formula (1),(2), the defective product probability will be calculated to the minimum and maximum values when passing through a particular facility in the total number of products. That is, if the defect rate for each equipment used in each process is calculated and constructed as a heat map, the figure 5 below is shown.

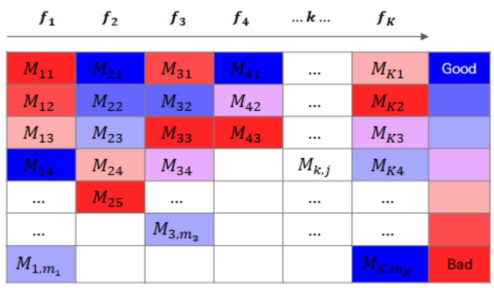

Figure 5:. The Visualization of facility

3.2. Naïve Bayesian

Naive Bayesian assumes the independence between properties by a classification technique, and defines the probability of belonging to c1 or c2 when a document d is given, and classified as a probability of belonging to it.[10] It is a classifier used mainly for document classification, and it is effective for next process scheduling because it has low complexity in calculation as in equation (3).

$$\mathrm{P}(c_1|d) = \frac{P(c_1,d)}{P(d)} = \frac{P(c_1,d)/P(c_1) * P(c_1)}{P(d)} = \frac{P(d|c_1)P(c_1)}{P(d)} \quad (3)$$

This study aims to develop the sequences of facilities to analyze data at manufacturing and determine follow-up processes through data mining technique analysis.

$$\mathrm{P}(\mathrm{Q} = \mathrm{B}|M_{k_1,j_1}, M_{k_2,j_2}) = \min_{j=1,2,3\ldots j} P(B)P(M_{k_1,j_1}|B)P(M_{k_2,j_2}|B) \quad (4)$$

The researchers determine whether the quality of the final product is high, average, or fail when the equipment is passed (condition), and if the product is classified as having a high probability value when passing through a particular facility, it is reflected in the scheduling decision.

The basic formula of the Naive Bayes is shown in equation (4),(5) and applied it to the assumed process to derive a new formula.

$$\mathrm{P}(C_k|\mathrm{x}) = \frac{P(C_k)P(x|C_k)}{P(x)} \quad (5)$$

In the example scenario, if the quality of the final product is high when passing through facilities $M_{1,1}$ and $M_{2,4}$, the following equation(6),(7) can be calculated.

$$\mathrm{P}(\mathrm{Q} = '\mathrm{high}'|f_1 = M_{1,1}, f_2 = M_{2,4}) = \frac{P('\square ig\square')P(M_{1,1}|'hig\square')P(M_{2,4}|'hig\square')}{evidence} \quad (6)$$

$$evidence = \mathrm{P}('\mathrm{high}')\mathrm{P}(M_{1,1}|'hig\square')P(M_{2,4}|'hig\square')$$
$$+P('average')P(M_{1,1}|'\mathrm{average}')\mathrm{P}(M_{2,4}|'average')$$
$$+P('fail')P(M_{1,1}|'fail')P(M_{2,4}|'fail') \quad (7)$$

Finally, the generalization of the equation is as following in equation (8),(9).

$$\mathrm{P}(\mathrm{Q}|f_1, f_2 \ldots f_k \ldots f_K) = \frac{P(Q)P(f_1|Q)P(f_2|Q)\ldots P(f_k|Q)\ldots P(f_K|Q)}{evidence} \quad (8)$$

$$evidence = \sum_{k=1,2\ldots K}^{Q=\square igh,average,fail} P(Q)P(f_1|Q)\ldots P(f_k|Q)\ldots P(f_K|Q) \quad (9)$$

4. The Suggestion of Algorithm and Further Analysis

This study processed the production history data of existing facilities and defined and constructed new data sets. In addition, we propose a production scheduling that can perform fast calculation, using the naive bayes and conditional probability analysis method, and can plan a scheduling with a high probability of producing high quality of final products, or low probability of low quality products.

In the basic statistical analysis of this paper, it was confirmed that the final quality of the product was changed following as the equipment number and the process which is the same.

However, it is still necessary to secure various actual production data as well as the algorithm proposal step. Therefore, we are going to establish a system that can automate the scheduling of facilities by performing the verification and optimization steps so that the algorithm can apply the actual process in the future.

Acknowledgment

"This research was supported by the MSIT(Ministry of Science, ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-2013-1-00881) supervised by the IITP(Institute for Information & communication Technology Promotion)"

References

[1] Shin Hyun-Joon, Ru Jae-Pil. (2010). An Adaptive Scheduling Algorithm for Manufacturing Process with Non-stationary Rework Probabilities, Korea Academy Industrial Cooperation Society, 11(11), 4174-4181

[2] So Byeong-Eob, Shin Sung-Sik. (2017). The Built of Smart Factory Using Sensors and Virtual Process Design, Journal of the KIECS, 12(6), 1071-1080

[3] T. Park, K. Park. (2012). Effect of IT Manufacturing Firms' Technological Innovation Factors - From Government Support Level, J. of Digital Convergence, 10(6), 17-22.

[4] Yield Management System. (2015). Retrieved from http://www.dbguide.net/knowledge.db?cmd=view&boardUid=186140&boardConfigUid=22&boardStep=&categoryUid=197.

[5] Application of Big Data in the Manufacturing Industry. (2016). Jo Seong-Jun, Shin Hun-Sik, Kim Seong-Hyeon.

[6] Jo Yong-Cheol, Jo Hyeon-Chan, Kim Jong-Won, Jang Ryang, Jeon Hong-Tae. (2008). A Study on Optimal Scheduling of Multi-Spinner's Manufacturing Process Using Artificial Neural Network, Journal of Korean institute of intelligent systems, 18(1), 157-160

[7] Kim Wook, Park Young-Moon. (1991). An Artificial Neural Network for the Optimal Path Planning, The proceedings of KIEE, 333-336.

[8] Ryu Sun-Joong. (2015). Development of Visual Inspection Process Adapting Naive Bayes Classifiers, JOURNAL OF THE KOREAN INSTITUTE OF GAS, 19(2), 45-53

[9] Semiconductor 8 Process 8th. First gate to acceptance EDS (Electrical Die Sorting) process. (2012). Retrieved from http://www.samsungsemiconstory.com/201

[10] Zhang, Harry. (2004). The Optimality of Naive Bayes, FLAIRS2004 conference.