Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp341-346)

Integration new Apriori algorithm MDNC and Six Sigma to Improve Array yield in the TFT-LCD Industry

Chiung-Fen Huang*, Ruey-Shun Chen**
* Institute of Information Management, Chiao Tung University
Management Building 2, 1001 Ta Hsueh Road, Hsinchu, Taiwan 300, ROC

Abstract: To increase process yield is the most effective way to raise income of TFT-LCD industry. This research is divided into two phases. In the first phase, We have modified Apriori algorithm called Multi-Dimension Non-Continuous (MDNC), an algorithm by eliminating the limitations imposed by traditional pattern matching of continuous data, to mine the association rules in the cross-day discrete manufacturing data and find out some valuable information.. The second phase use of MDNC system integrated with six-sigma quality project is set up as the assessment platform which indexes performance and improve performance tracing, to make sure the outcomes of project is profitable enough to reward financial benefit. The preliminary results show the system of new Apriori algorithm techniques integrated with six-sigma methodology is proposed to improve process yield of the Array manufacturing process, the deforming rate of production lines, reduce the cycle time, increase the profit and make the CpK value stable.

Key-Words: Apriori Algorithm, Data Mining, TFT-LCD, *six-sigma*

1 Introduction

If the information of manufacturing process can be applied correctly, the process yield will be raised and so will the competitiveness of enterprise. To achieve the above aims, in the paper, the system of data mining techniques integrated with six-sigma should be practiced.

Past traditional improvement techniques such as formulation of process yield and simulation must depend on sophisticated statistical application and skills. However, the complex interactions of operational parameters and nonlinear characteristics of manufacturing process make the above traditional methods constrained. Furthermore, it is impracticable to depend on manual labor to acquire the information of decision making from a large number of data in the database of LCD manufacturing process. Data mining is one of the most effective methods which can induce the useful information or rules to improve the process yield from lots of the original rough data sets.

In the framework of this research, the fitness of new Apriori Algorithm techniques integrated with six-sigma project to be applied in the front process of Array manufacturing will be discussed. This paper is expected to achieve the following objectives: significant improvement of process quality, reduction of production cost, increasing the income of enterprise, achievement of higher delivery rate, and the enhancement of enterprise competitiveness.

2 data mining and MDNC algorithm

2.1 Association Rule Mining

Association rule mining is used to search for interesting relationships among data items in a given data set. The first algorithm of mining association rules was proposed by Agrawal, Imeilinski, and Swami in 1993[6].

Association rule can be analyzed for buying patterns that reflect items frequently associated or purchased together. For example, the information that customers who purchase computers also tend to buy financial management software at the same time is represented in Association Rule (1) below[7]:

Computer → financial_software
[support = 20%, confidence = 50%] (1)

A support of 20% for Association rule (1) means that 20% of all the transactions under analysis show that computer and software are purchased together. A confidence of 50% means that 50% of the customers who purchased a computer also bought the software.

Let I = { i1, i2, …, im } be a set of items. Let D, the task-relevant data, be a set of database transactions where each transaction T is a set of items such that

By the definition of confidence and support:

$$confidence(A \Rightarrow B) = \frac{\#\_tuples\_containing\_both\_A\_and\_B}{\#\_tuples\_containing\_A} \quad (2)$$

$$\text{support}(A \Rightarrow B) = \frac{\#\_tuples\_containing\_both\_A\_and\_B}{total\_\#\_tuples} \quad (3)$$

Rules that satisfy both a minimum support threshold and a minimum confidence threshold are called strong.

2.2 Modified Apriori Algorithm

The first algorithm of mining association rules, called Apriori algorithm, was proposed by Agrawal and Srikant in 1994[8]. The methodology is so important in frequent pattern mining that there are many approaches adapting from it.

In order to fit our experimental model, we modify Apriori algorithm as described further below.

Fig.1 shows the traditional Apriori algorithm. Next, MDNC algorithm allows users to define the maximum days for cross-day transaction, and successively generates various cross-day data according to the maximum setting days (Fig.2). After saving the assorted data through an array, the formula can automatically locate all possible correlations during the appointed maximum days (Fig.3). In the end, the min-support threshold and min-confidence threshold are set up to discover the relationship (Fig. 4). The result could be a particular model that exists only on the mth or nth day, consecutively or not. In applying this discrete proposal, the limitations imposed by pattern matching of continuous data are eliminated, and more helpful information is revealed.

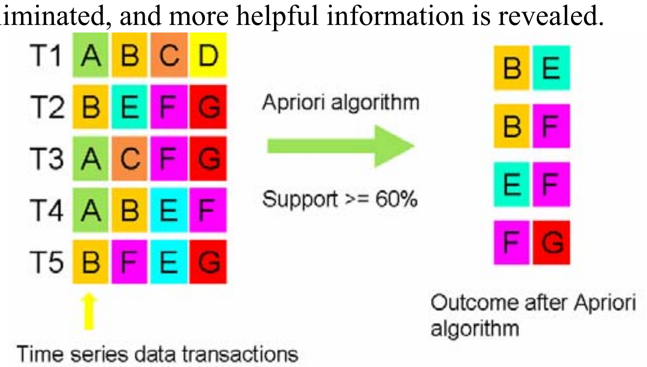

Fig.1 Traditional Apriori algorithm example

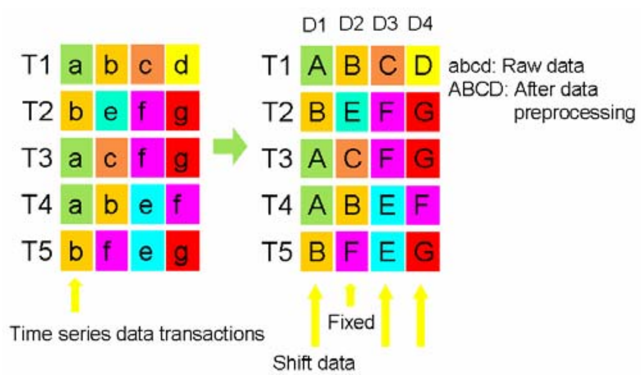

Fig.2 MDNC algorithm (Data preprocessing)

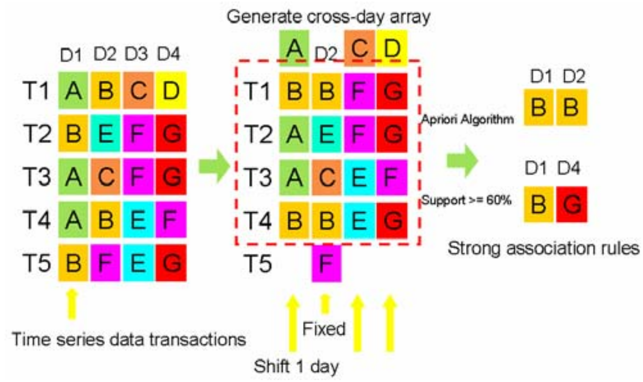

Fig.3 MDNC algorithm (shift 1 day)

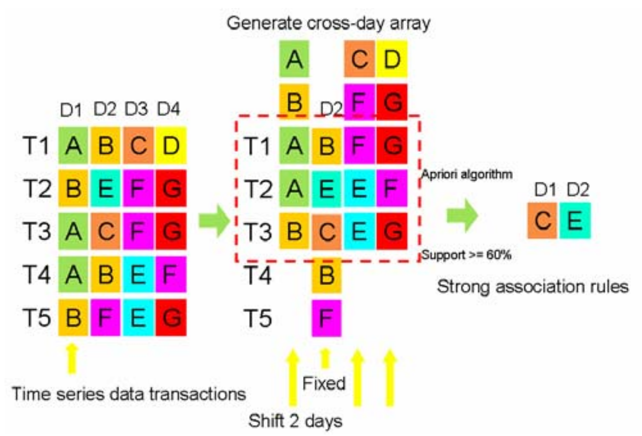

Fig.4 MDNC algorithm (shift 2 days)

Although a data mining process may generate a large number of patterns, typically only a small fraction of these patterns will actually be of interest to the given user. In this paper, we use novelty test measures of pattern interestingness.

3 Experiment

3.1 A data mining system

The sources of data are from Chunghwa Picture Tubes Inc.(CPT), one of TFT-LCD panel makers in Taiwan. Manufacturing data were collected and experimented from June 16, 2003 to Sep. 15, 2003. The raw data include production quality information and processing history of each lot as shown in Table 1

Table 1 Example Of Production Quality Info.

Output date	LOTID	Test input	Test output	SO	GO	GG	SS	GS	GC	SC
2003/6/16	FJ1JS0	80	76	0	1	0	3	0	0	0
2003/6/17	FJ1KF0	76	67	4	1	2	0	0	1	1
2003/6/16	FJ1MG0	80	66	5	1	1	1	3	3	0
2003/6/16	FJ1PF0	80	73	4	0	0	2	0	1	0
2003/6/16	FL8NN0	80	73	4	1	0	1	0	1	0
2003/6/17	FX1JP0	80	72	2	1	2	0	0	2	1
2003/6/17	FX1KY0	80	70	4	1	1	1	2	1	0
2003/6/17	FX1LU0	80	75	1	1	0	3	0	0	0
2003/6/18	FX1M00	76	71	3	1	0	1	0	0	0
2003/6/16	FX1MD0	80	75	1	2	1	0	1	0	0
2003/6/17	FX1MQ0	80	73	3	1	0	2	0	1	0
2003/6/18	FX1NL0	80	74	1	1	1	2	0	1	0
2003/6/16	FX1NN0	80	73	4	0	1	1	0	1	0

Step 1: Date preprocessing

Transform and calculate the amount of defect for each line defect type from the raw data of manufacturing database. For example, Table 2 shows that the data below 90% yield rate are selected and calculated for line defect S-Open.

Step 2: Two items are selected from machine group, mask type, or machine No. as the variables and the rest are kept as constants to find the combination with higher defect rate.

We can get:

Support (machine group 3 → mask type 2) = 0.76%

Confidence (machine group 3 → mask type 2) = 2.79%

Step 3: Machine group, mask type, and machine No. are selected as the variables to find the combination with higher defect rate.

As the Table 3, combination (machine group 1, mask type 3, and machine no. 1) shows the highest defect rate.

$\forall\ x \in T,\ machine\ group(X,1) \wedge mask\ type(X,3) \Rightarrow machine\ no.(X,1)\ [32.4\%,56.4\%c]$

Step 4: The Support and Confidence from Step 2 to 3 are compared. The combination with the maximum value might cause the highest defect rate.

We can find mask type 3 with machine No.1 in the first machine group should be first priority to stop production to avoid products defect.

Step 5: Summary

We can base on procedure step1 to setp4 to find out all of problematic machines that cause line defect and summarize as Table 4.

Table 2 Defect Distribution for S-Open

Machine group	Mask type	Machine No.	Num. of defect
1	3	1	254
1	3	2	7
1	3	3	189
2	2	5	90
2	4	4	25
3	2	4	6
3	3	1	14
3	3	2	164
3	3	8	20
3	4	2	5
3	4	5	6
6	3	2	4

Table 3 Strong Association Related To Defect

machine group	Mask type	Machine No.	Num. of defect	Supp.	Conf.
1	3	1	254	32.40%	56.40%
1	3	2	7	0.90%	1.60%
1	3	3	189	24.10%	42.00%
2	2	5	90	11.50%	78.30%
2	4	4	25	3.20%	21.70%
3	2	4	6	0.80%	2.80%
3	3	1	14	1.80%	6.50%
3	3	2	164	20.90%	76.30%
3	3	8	20	2.60%	9.30%
3	4	2	5	0.60%	2.30%
3	4	5	6	0.80%	2.80%
6	3	2	4	0.50%	100.00%

Table 4 Analysis of Machine Defect Rate

PVD	GE	SE	SD	CH	PE	Summary
ASA010	0.00%		2.54%		0.00%	2.54%
ASH010	1.19%		1.40%		1.77%	4.36%
ASM010	1.66%		7.43%		0.00%	9.08%
ASM020	2.43%		5.33%		0.00%	7.77%
ASP010	0.00%		0.00%		1.58%	1.58%
PHO	GE	SE	SD	CH	PE	Summary
APC010	0.00%	0.00%	2.54%	0.10%	0.00%	2.64%
APC020	0.00%	0.00%	0.82%	0.29%	0.00%	1.11%
APC030	0.00%	0.18%	0.00%	0.00%	0.00%	0.18%
APC040	0.00%	0.07%	0.00%	0.00%	0.00%	0.07%
APC050	0.00%	0.00%	0.00%	0.42%	0.37%	0.79%
APC060	0.52%	0.00%	3.44%	0.00%	0.00%	3.95%
APC070	1.20%	0.00%	0.00%	0.00%	0.00%	1.20%
APC080	0.00%	0.00%	0.97%	0.06%	0.00%	1.03%
APC090	0.00%	0.00%	0.00%	0.00%	0.29%	0.29%
CVD	Gi	SE	SD	CH	PE	Summary
AVB010	1.25%	1.06%		0.66%		2.97%
AVB020	0.81%	0.92%		0.12%		1.84%
AVB030	2.19%	1.56%		0.37%		4.13%
AVB040	7.41%	1.25%		0.42%		9.08%
AVB050	0.78%	0.65%		0.50%		1.92%

3.2 Execution procedure of data mining system integrated with six-sigma

Six Sigma project is divided into 5 phase as follows: define (D), measure (M), analyze (A), improve (I) and control(C). It is called DMAIC [9],[10].

Phase 1 – define: Before the six-sigma project starts to be executed, the data collected from June of 2003 to August of 2003 should be finished by statistical analysis. Based on the data in the table 5, the total amount of the tested TFT templates is 93093 pieces. There are 3171 pieces of TFT template which has the defect of SO or the defect of SS among all of the TFT templates, so the defect rate is about 3.41%. In the aspect of attribute analysis of defect causes, the defects rate caused by PVD equipments is 48.51%, in other words, 1.65% out of 3.41% defect rate is caused by PVD equipments(Table 5). So the final goal of the six-sigma project is to improve the line defects (SO, SS) and point defects caused by PVD equipments.

Table 5 The rate of root cause of Line defects form PVD process

Week	Test panels	Line defect SO+SS	SO+SS (%)	Root cause by PVD (%)	SO+SS by PVD (%)
2003/6/30	5812	192	3.30%	54.15%	1.79%
2003/7/7	11948	390	3.26%	49.73%	1.62%
2003/7/14	12368	453	3.66%	48.09%	1.76%
2003/7/21	14881	449	3.02%	47.92%	1.45%
2003/7/28	22656	721	3.18%	53.79%	1.71%
2003/8/4	11072	458	4.14%	51.49%	2.13%
2003/8/11	5728	226	3.95%	33.60%	1.33%
2003/8/18	8628	282	3.27%	49.31%	1.61%
Total	93093	3171	3.41%	48.51%	1.65%

Phase 2 – measure: Find the key factors of PVD manufacturing process which may affect CTQ.

The PVD equipments are used in the manufacturing process of both Mask1 and Mask3. Mask1 is the manufacturing process of GE, and Mask3 is the manufacturing process of SD. Refer to the analysis information from the table 6, the percentage of bad TFT templates found in the manufacturing process of Mask3 is from 40% to 50%, the percentage of bad TFT templates found in the manufacturing process of Mask1 is below 10%. As the fig.5 shows, the X-axis represents the magnitude of defect rate, and the left values are smaller than their values of right side, so the central points of the normalized distribution diagram for the manufacturing process of GE and SD will be compared. The SD ones is more right than the GE ones, so it means the defect rate of the manufacturing process of SD is higher than the one in the manufacturing process of GE.

Table 6 Root cause analysis data of PVD process

Week	Quantity of sample	L/d cause of PVD process in Mask1		L/d root cause of PVD process in Mask3	
		Qty.	(%)	Qty.	(%)
2003/6/30	215	4	1.86%	107	49.77%
2003/7/7	195	7	3.59%	84	43.08%
2003/7/14	136	9	6.62%	54	39.71%
2003/7/21	308	15	4.87%	123	39.94%
2003/7/28	289	9	3.11%	140	48.44%
2003/8/4	106	8	7.55%	44	41.51%
2003/8/11	132	11	8.33%	31	23.48%

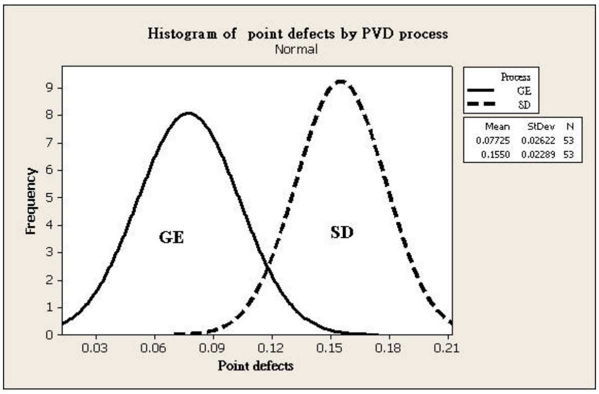

Fig.5 Histogram of point defect rate by PVD

Phase 3 – analyze: The most important defect cause in SD manufacturing process is because of a great number of particles.

Phase 4 – improve: Each method has its execution time and its own combination of equipments, so the use of the data mining system will divide the results generated from the different combination of equipments. Afterward, the data of results will be analyzed by Minitab. The experiment results of the line defects of SO and SS are respectively shown in the fig.6 and fig.7. The fig.6 is X bar-R Chart. The above of the fig.6 shows the tracing of the average values of the manufacturing process, and the below one shows the tracing of the variations of the manufacturing process. A1, A2, B1, and B2 denote the four methods. The average values, the variations and the discrete degree based on the method B2 are all the smallest in the fig.6. The fig.7 is the normalized distribution diagram. Compare the central values and the scatter range in the figure, if the central values go more left, then defect rate will decrease. So it is obviously that using the method B2 is better than using the method A2.

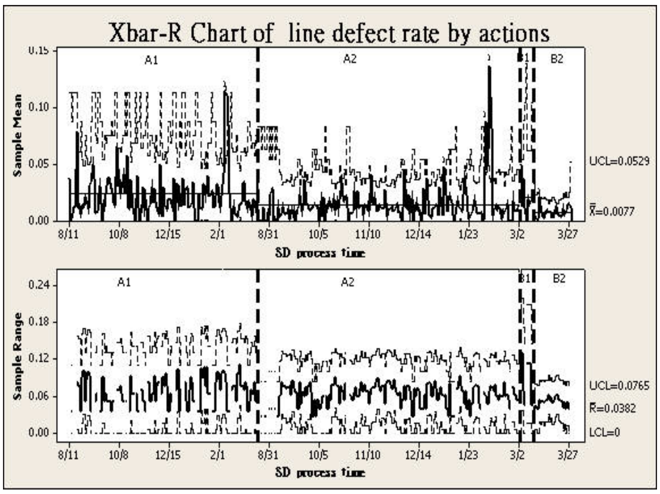

Fig.6 Line defect rate by action

Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China, April 16-18, 2006 (pp341-346)

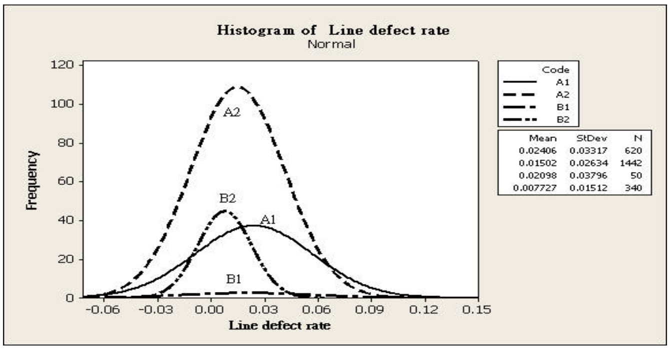

Fig.7 Histogram of Line defect rate by action

Phase 5 – control: Correct the condition of the manufacturing process, and implement the method B2. Furthermore, the above figures and tables of data mining can be treated as the tracing index of quality improvement performance.

4 Beneficial analysis and feedback

The system in the paper provides a good platform to run six-sigma project. The schedule of the data mining system integrated with six-sigma project is shown as follows:

a. 2003/06/16 ~ 2003/09/15: The data mining system is just on trial. Test the data mining system and do some debugging job. The process yield rate is improved from 90.7% to 95.3%. Set the base line for the six-sigma project by Analyzing the causes that make the defect rate 4.7%.

b. 2003/09/20 ~ 2004/03/31: Start to run the data mining system and do research on some specified quality problems in the six-sigma project. Apply techniques of DMAIC and the statistical tool to testify the feasibility of every project. And finally find out the feasible plans for the manufacturing process.

c. 2004/04/01 ~ 2005/03/31: Trace the executive performance of the research result transferring to the production lines. Proof the project has achieved the anticipated targets.

The results of the entire beneficial analysis are listed as the followings:

4.1 Performance of project

After implementing the six-sigma project, the percentages of the line defects are under the Base line at 1.46%. The improvement performance on decreasing the percentage of the line defects is shown as the Fig.8. The average improving rate for the defects is close to 0.74%. The degree of quality improvement is up to 50%. The financial benefit is also shown in the Table 7; the net profit can be raised about 0.76% in average.

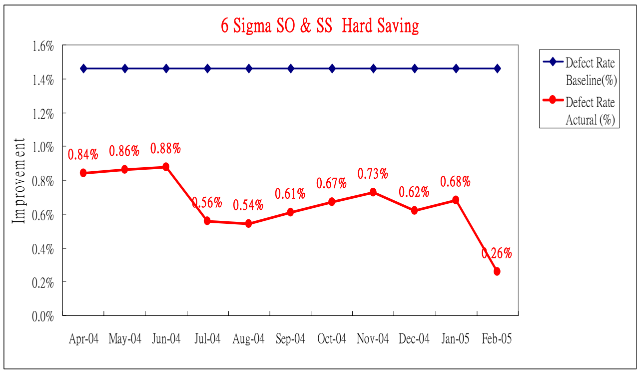

Fig.8 The monthly performance assessment of quality improvement on the line defects

Table 7 The table of financial benefit

Month/Year	04/2004	05/2004	06/2004	07/2004	08/2004	09/2004
Actural(NT$ 10K)	348.5	440.9	286.4	278.6	390.9	467.1
Target(NT$ 10K)	462.9	462.9	462.9	462.9	462.9	462.9
Yield%	75.3%	95.2%	61.9%	60.2%	84.4%	100.9%
Month/Year	10/2004	11/2004	12/2004	01/2005	02/2005	03/2005
Actural(NT$ 10K)	354.2	314.8	369.3	344.3	328.1	526.3
Target(NT$ 10K)	462.9	462.9	462.9	462.9	462.9	462.9
Yield%	76.5%	68.0%	79.8%	74.4%	70.9%	113.7%
Total	(NT$ 10K)	4449.4				

4.2 Index of side effect---Lot Hold and Delivery Date

The length of average delivery time is another index to observe the side effects. The data is shown as the table 8. The trail period of data mining system of CIM is 9.2 days. In the supervisory of the project in 2004, the average delivery time for normal Lot is about 8.5 days. According to the above conditions, there is no side effect in the execution period of the project.

Table 8 The list of average period of Lot

Month	Normal Lot			14 days~20 days			over 21 days			working period	
	Quantity	Weighting	Average	Quantity	Weighting	Average	Quantity	Weighting	Average	Average working period	Profit gain over working period
Jun-03	288	64.9%	10.4	72	16.2%	17.4	84	18.9%	37.5	16.6	-
Jul-03	1078	86.5%	9.8	107	8.6%	17.0	61	4.9%	28.9	11.4	31.61%
Aug-03	1335	93.9%	8.6	38	2.7%	16.9	49	3.4%	29.3	9.5	42.80%
Sep-03	1226	88.9%	9.1	110	8.0%	16.8	43	3.1%	24.9	10.2	38.71%
Oct-03	977	96.2%	8.7	28	2.8%	16.7	11	1.1%	21.9	9.0	45.57%
Apr-04	1071	76.7%	9.3	245	17.5%	16.7	81	5.8%	35.9	12.2	26.73%
May-04	1410	90.8%	8.9	114	7.3%	16.6	29	1.9%	33.3	9.9	40.3%
Jun-04	1662	95.0%	8.1	56	3.2%	15.6	31	1.8%	49.8	9.1	45.1%
Jul-04	1774	94.3%	8.2	60	3.2%	16.1	48	2.6%	36.8	9.1	44.9%
Aug-04	1324	91.3%	8.8	82	5.7%	16.5	44	3.0%	55.5	10.6	36.0%
Sep-04	1479	95.3%	7.5	49	3.2%	16.7	24	1.5%	38.2	8.3	50.3%

4.3 Discussion of Cpk values of average delivery date

To assess the benefit of the data mining system of CIM in the six-sigma project the points of view of process capability is used. The fig.9 is the achievements check table of after implementing the six-sigma project. During the project execution period of 2004, stricter standard is adopted in calculation of the average delivery rates. The upper limit of delivery date is 16 days based on the data from the table 8. About 95% of the Lots are done within 16~17 days. Compare the 21-day delivery date of production lines in 2003; there is a big progress during that time and the variation of the delivery time is less and steady, it is always about 2 days.

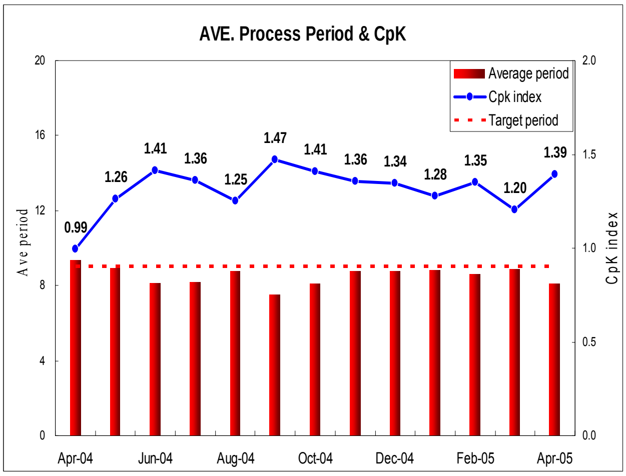

Fig.9 Average period and Cpk index in 2004

5 Conclusion

The contribution of this paper is to provide a system which can find the association rules from the data of Array manufacturing process to correct the combination of equipments and increase the process yield as well and enhance the competitiveness of enterprise. The research results are as the followings:

˜ Before this data mining system is implemented, the average process yield rate was 90.7%. The average process yield rate has increased 4.6% to 95.3% after the system was implemented

˜ After six-sigma is used in the project, the performance of quality improvement has been increased. The defect proportion of single cause decreased from 1.42% to 0.72%, which means the total yield rate, has been increased to 5.34%.

Concerning about the financial benefit of the project, the incomes of enterprise can be increased 0.76%. Through the experiment in the paper, therefore, it is workable to apply the data mining techniques integrated with six-sigma project on the real manufacturing process. On the other hand, the above techniques can help the execution of procedure improvement project stay in a stable manufacturing process environment, improve the process yield and increase the incomes of enterprise.

References:

[1] S. P. Cunningham, C. J. Spanos, and K. Voros, Semiconductor yield improvement: Results and best practices, *IEEE Trans. Semiconduct. Manufact*., May 1995,vol. 8, pp. 103-109

[2] W. Kuo and T. Kim, An overview of manufacturing yield and reliability modeling for semiconductor products, *Proc. IEEE*, Aug. 1999,vol. 87, pp. 1329-1344

[3] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, From data mining to knowledge discovery: An overview, *in Advances in Knowledge Discovey and Data Mining*, MIT Press,1996, pp. 1-36.

[4] M. J. Berry and G. Linoff, *Data Mining Techniques*. New York:Wiley, 1997.

[5] R. Groth, *Data mining: building competitive advantage*, Prentice-Hall, Inc., 2000.

[6] R. Agrawal, T. Imielinski and A. Swami, Mining association rules between sets of items in large databases, *in Proceedings of ACM Special Interest Group on Management of Data*, 1993, pp.207-216

[7] J. Han, and M. Kamber, *Data mining concepts and techniques*, Morgan Kaufmann, San Francisco, 2000.

[8] R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large database, *Proc. 20th Int'l Conf. On very Larg Databases*, Santiage, Chile, Sep. 1994, pp. 478-499

[9] Hendericks, C. and Kelbaugh, R. (1998), Implementing Six Sigma at GE, *The Journal for Quality and Participation*, 1998.

[10] Kim M. Henderson and James R. Evans, Successful implementation of Six Sigma: benchmarking General Electric Company, *Benchmarking-An International Journal*, Vol. 7 No. 4, 2000, pp. 260-281.