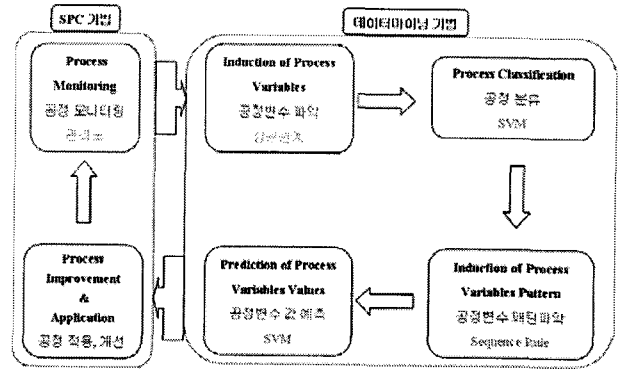




터마이닝기법은 데이터들이 복잡하게 영향을 미치는 분야에 적합하고 반도체 클리닝공정을 온라인으로 모니터링하기에 효율적이라고 제안하였다[4]. 백동현등은 데이터마이닝기법을 이용하여 반도체 FAB공정의 불량률을 판별하고, 불량원인 장비 및 공정변수 파악을 통하여 반도체의 수율을 예측하는 정보시스템을 제안하였다[3]. 그러나 기존의 연구들은 데이터마이닝을 통하여 공정에 영향을 주는 변수파악 및 예측에 중점을 두었으며 통계적 공정관리기법과의 접점은 미흡하였다. 본 연구에서는 통계적 공정관리기법을 통하여 공정의 상태를 파악하고 공정의 이상 발생시 공정으로부터 수집된 데이터를 데이터마이닝기법을 통하여 공정의 불량에 영향을 미치는 공정변수의 분류, 패턴파악, 예측을 통하여 공정을 개선하고 이를 다시 통계적 공정관리기법으로 피드백하여 공정을 안정상태로 유지관리 하는 관리기법을 제시하고자 한다.

으로 피드백 하여 공정에 적용 및 개선하는 공정 적용 · 개선단계를 거치므로 인하여 관리도를 수정을 하여 공정을 안정상태로 유지, 관리할 수 있도록 한다. 공정 변수 파악 및 개선방법의 제안모형은 다음 <그림 1>과 같다.



<그림 1> 제안 모형

## 2. 공정변수 파악 및 개선방법

### 2.1 제안 모형

통계적 공정관리기법인 관리도는 공정을 모니터링하고 공정의 이상을 파악하기가 용이하다. 그러나 현재 기업의 공정은 복잡해지고 공정변수가 많아짐으로서 관리도만을 적용하여 공정의 이상원인을 파악하기 힘들다. 그러므로 본 연구에서는 대용량의 데이터로부터 정보를 찾아 문제를 해결하는 방법인 데이터마이닝기법과 통계적 공정관리기법을 적용하여 공정에 불량요인 공정변수 파악 및 공정 개선하는 모델을 제시하고자한다. 본 연구에서 제시하는 단계를 살펴보면 다음과 같다. 공정 모니터링 단계에서는 관리도를 이용하여 공정을 모니터링 한다. 만약 공정의 이상이 발생할 경우 데이터마이닝기법을 적용한다. 데이터마이닝기법을 적용하기 전 사전처리로 공정과악 단계에서는 공정 흐름도(process map)를 통하여 공정에 영향을 주는 공정변수를 파악한다. 파악된 공정변수를 통하여 공정 분류단계에서 목표불량률보다 불량률이 높게 나타난 공정패턴과 낮게 나타난 공정패턴을 SVM(Support Vector Machines)기법으로 공정조건을 분류한다. 분류된 공정패턴에서 불량률이 높게 나타난 공정패턴에 순차적 연관규칙(sequence rule)을 적용하여 불량률에 영향을 많이 주는 공정변수의 패턴을 파악하는 공정변수 패턴파악단계를 거치고 공정변수 값 예측단계에서 SVM기법을 이용하여 공정변수 패턴파악단계에서 파악된 공정변수를 안정된 공정의 공정변수 값으로 예측을 하는 단계이다. 예측된 공정변수 값을 공정

### 2.2 공정 모니터링 (Process Monitoring)

공정을 모니터링하기 위하여 관리도를 많이 적용하고 있다. 관리도를 공정에 적용할 경우에는 관리도의 목적에 맞는 관리도를 사용하여야 한다. 왜냐하면 공정에 부합되지 않은 항목에 대하여 관리도를 사용하면 공정을 효과적으로 관리할 수 없기 때문이다. 관리도가 결정되면 관리도의 설계모수를 설정하여야 한다.

설계모수로는 샘플크기, 샘플링 간격, 관리한계선등이 있다. 관리도의 설계모수가 결정되면 공정을 모니터링한다.

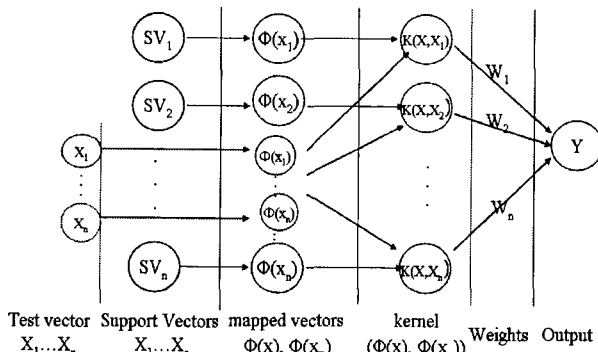
### 2.3 공정변수 파악 (Induction of Process Variables)

공정의 이상원인 발생시 공정변수의 데이터를 수집하여 분석해야 한다. 해당공정을 파악하기 위하여 공정의 흐름도를 작성하고 공정변수의 데이터를 수집한다. 수집된 공정변수 데이터들이 공정에 얼마나 영향을 미치는지를 파악하여 주요 공정변수를 선별한다.

### 2.4 공정 분류 (Process Classification)

공정의 상태를 불안정과 안정한 공정으로 분류하는 것으로 공정의 목표불량률에 의하여 공정을 2진 분류한다. 데이터마이닝 기법에서의 분류방법은 SVM기법, MLP기법, 의사결정 나무기법 등이 많이 사용되고 있다. 본 연구에서는 SVM기법을 이용하여 분류를 실시하였다.

SVM기법은 Vapnik에 의하여 제안되었으나 컴퓨터 기술의 미흡으로 적용되지 못하다가 최근 컴퓨터기술의 발달로 각광받게 되었다. 또한 Vapnik(1995)는 패턴인식이나 비선형회귀모형 등 효과적으로 수행할 수 있는 방안을 제시하였다[5][11]. LeCun et al.은 SVM기법이 MLP(Multilayer Perceptron)기법 보다 뛰어난 성능을 보임을 확인하였다[7]. SVM기법과 기존의 통계적 학습 방법과의 다른 점은 기존의 경험적 위험 최소화를 이용하는 방법이 아닌 구조적 위험 최소화를 이용하여 오차를 줄이는 방법을 사용하는 것이다. <그림 2>는 본 연구의 SVM기법의 구조를 나타낸다.



<그림 2> SVM기법의 구조

2.5 공정변수 패턴파악  
(Induction of Process Variables Pattern)

분류된 공정조건에서 불량률이 높은 공정조건을 통하여 공정변수들의 패턴을 파악하는 단계이다. 본 연구에서는 데이터마이닝 기법 중 Zaki에 의해 제안된 순차적 패턴규칙(sequence pattern rule)을 적용하였다. 순차적 패턴규칙은 일반 연관규칙과 달리 어떤 사건이 시간의 순서에 의하여 발생할 경우 적용하는 기법이다[12]. 일반적인 생산 공정은 일정한 순서에 의하여 공정이 진행 되므로 순차적 연관규칙을 이용하여 공정의 패턴을 찾을 수가 있다. 순차적 패턴규칙의 측도는 지지도와 신뢰도이다.

2.6 공정변수 값 예측  
(Prediction of Process Variables Values)

공정에 영향을 주는 불안정 공정변수를 안정된 공정변수 값으로 예측한다. 공정변수패턴 파악을 통한 불안정한 공정변수 값을 공정이 안정한 상태를 유지할 수 있는 공정변수 값을 제시한다. 본 연구는 예측을 위한 기법으로 SVM기법을 적용하였다.

2.7 공정적용 및 개선  
(Process Improvement & Application)

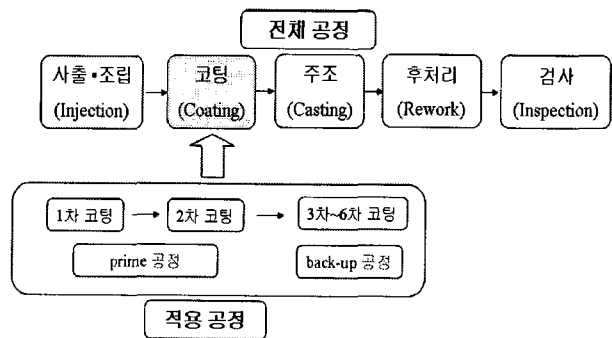
예측된 공정변수 값을 공정에 적용하여 관리도를 수정함으로써 공정의 안정화를 수립하고 지속적으로 공정을 모니터링한다.

3. 사례연구

본 연구의 모델을 적용시키기 위하여 정밀주조를 이용하여 부품을 생산하는 국내 K기업의 코팅공정 데이터를 활용하였다.

3.1 공정소개

정밀주조의 공정은 모형을 사출하여 몰드(mold)에 조립하는 공정과 조립한 몰드에 코팅을 하는 코팅공정, 급속(ingot)을 녹여 몰드에 붓는 주조공정, 주조된 몰드를 절단, 연마를 하는 후처리공정, 최종제품을 검사하는 검사공정으로 나눌 수 있으며 본 연구에서는 코팅공정에 제안하는 모형을 적용하였다. 코팅공정은 코팅액(slurry)의 차수에 의해 공정이 세분되어진다. 코팅의 차수는 1차 와 2차코팅을 'primely 공정'이라고 칭하고 나머지 3차부터 6차코팅까지 공정은 'back-up 공정'이라고 한다. 이 기업의 전체 생산 공정과 적용공정의 흐름도는 다음 <그림 3>과 같다.

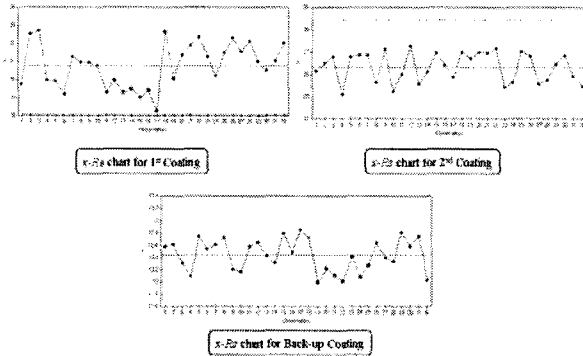


<그림 3> 전체공정과 적용공정 흐름도

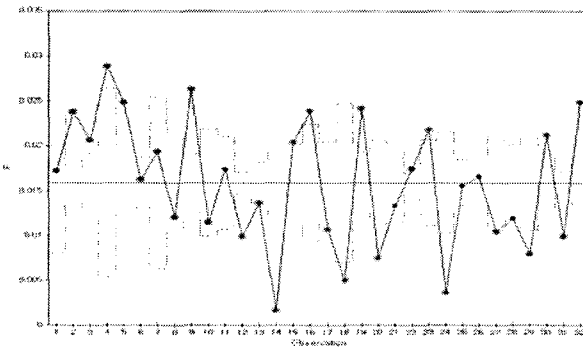
3.2 공정 모니터링

코팅공정의 세부공정을  $x-Rs$  관리도로 모니터링을 한 결과 공정의 이상이 파악되지 않았다. 그러나  $p$  관리도를 통하여 전체 공정에서의 코팅공정에 불량률을

모니터링 한 결과 공정의 이상이 파악되었다. 그러므로 코팅공정의 공정변수에 대한 공정관리 조건이 잘못 설정되었다고 판단된다. <그림 4, 5>에서는 공정의 모니터링 된 관리도를 나타낸 것이다.



<그림 4>  $x - R_s$  관리도를 통한 공정 모니터링



<그림 5>  $p$  관리도 통한 공정 모니터링

3.3 공정변수 파악

코팅공정의 이상원인을 분석위해 상관관계분석을 실시하였으며 척도로는 피어슨(Pearson's)의 R계수를 이용하였다. 분석 결과는 <표 1>에서 나타난다.

<표 1> 상관관계

	1 차 점도	2 차 점도	백업 점도	온도	습도
불량률 p-value	0.39165 <.001	0.24373 <.0001	0.06324 <.032	0.08 0.395	-0.01 0.834

코팅공정의 각 차수의 점도는 Zancup #4로 측정되며 측정단위는 초이다. <표 1>의 분석 결과 각 차수의 점도

가 유의수준 0.05에 대하여 불량률에 유의한 결과가 나타났다. 그리고 공정의 온도, 습도는 항온항습기에 의하여 관리하기 때문에 값이 유의하기 않은 결과가 나타났다.

3.4 공정분류

공정의 불량률 데이터를 이용하여 공정을 2진 분류를 하였다. 데이터는 학습용 데이터로 200개, 평가용 데이터로 50개를 사용하였다. 본 연구에서 적용한 기법인 SVM기법[6]과 MLP기법[1]을 적용 분류결과는 <표 2>와 같다.

<표 2> SVM기법과 MLP기법의 오분류율 비교

분류기법	오분류율
SVM	0.26%
MLP	0.36%

3.5 공정변수 패턴파악

공정분류를 통한 데이터를 이용하여 공정의 이상원인을 주는 공정변수 패턴을 파악하기 위하여 공정변수를 <표 3>과 같이 변환하였다.

<표 3> 공정변수 변환

변 수			변수명
하한	중심	상한	
A- (30초 이하)	A (30~31초)	A+ (31초 이상)	1차 점도
B- (24초 이하)	B (25~27초)	B+ (27초 이상)	2차 점도
C- (12초 이하)	C (12~13초)	C+ (13초 이상)	백업 점도

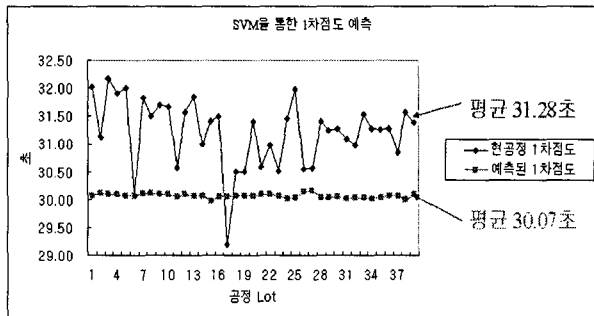
변환된 공정변수를 순차적 패턴규칙을 적용하여 불량률에 영향을 미치는 공정변수의 패턴을 분석한 결과 가장 많이 영향을 미치는 패턴은 <표 4>와 같다. 즉, A(1차점도)가 높게 나타났을 경우 불량률이 많이 발생하였으며 이는 몰드의 shell 불량률이 발생함을 알 수 있다.

<표 4> 순차적 패턴규칙 적용 결과

지지도(%)	신뢰도(%)	공정 변수의 패턴	불량 형태
75.86%	95.65%	A+ → B → C	shell 불량 (shell 더짐, shell 크랙)
A+: 1차 점도 상한 B: 2차 점도 중심 C: 백업 점도 중심			

3.6 공정변수 값 예측

코팅공정의 1차점도 값이 공정의 불량에 많은 영향을 미친다는 분석결과를 통하여 1차공정의 공정변수 값을 SVM기법을 통하여 안정된 1차점도를 예측하였다. <그림 6>은 1차 점도에 대한 현 공정과 예측된 공정변수의 값이다.



<그림 6> 현 공정과 예측된 공정변수 값 비교

예측에 사용된 학습용 데이터는 200개, 평가용 데이터는 40개가 분석에 활용하였다. <그림 6>에서 나타난 결과 현 1차 공정의 공정평균의 점도는 31.28초로 나타났으며 안정된 공정으로의 예측된 1차 공정의 점도는 30.07초로 나타났다. 그러므로 현 공정의 점도가 높게 공정이 진행되고 있음을 알 수 있으며 이로 인하여 공정의 불량이 높게 나타남을 알 수 있다.

3.7 공정개선 및 적용

현 공정의 공정불량 형태는 shell 터짐과 shell 크랙이며 이 불량률의 원인은 shell 강도가 높아서 생기는 불량이다. 이는 shell 강도와 통기도 테스트를 실시하여 파악할 수 있다[정밀주조소재 제조작업표준, 1981]. 그러므로 안정된 공정으로 예측된 평균점도 30초와 현 공정의 평균점도 31초를 shell강도/통기도 테스트 실시하였으며 결과는 <표 5>와 같다.

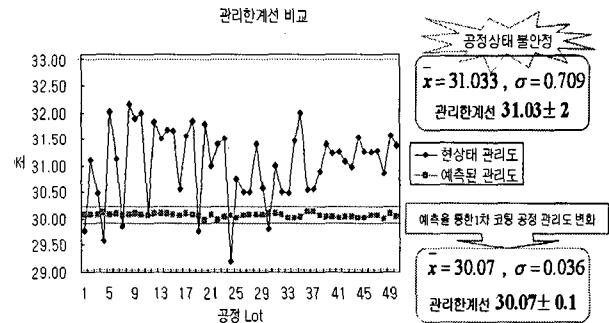
<표 5> shell 강도/통기도 테스트 결과

	shell 강도 test		shell 통기도 test	
	예측된 1차 점도	현 공정의 1차 점도	예측된 1차 점도	현 공정의 1차 점도
규 격	30 ~ 60 kgf/cm <sup>2</sup>		100 ~ 160 cc/min	
평 균	38.85 kgf/cm <sup>2</sup>	47.18 kgf/cm <sup>2</sup>	133.75 cc/min	126.25cc/min
표준편차	4.382	4.055	2.5	2.5
샘플수	4	4	4	4

<표 6> 통계적 검정 결과

	shell 강도	shell 통기도
가설	제한하는 점도와 현 공정의 점도를 통한 Shell 강도 값이 같다	제한하는 점도와 현 공정의 점도를 통한 Shell 통기도 값이 같다
검정 통계량	-2.79	5.567
p-value (95%)	0.016	0.0007
검정결과	가설 기각	가설 기각

테스트 결과를 통계적 검정을 통하여 분석한 결과는 <표 6>과 같이 나타났으며 검정을 통한 결과는 예측을 통한 1차점도 값과 현 점도와의 차이가 있음을 알 수 있고 shell 강도는 낮아졌고 통기도는 높아졌음을 확인할 수 있다. 그러므로 현 평균점도 31초에서 예측된 평균 30초로 공정의 관리한계선을 수정하여야 한다. <그림 7>에서 50개의 공정데이터를 예측을 통하여 관리한계선을 수정한 관리도이다. 그 결과 예측을 통한 관리한계선의 σ 값이 낮아짐을 통하여 공정이 안정됨을 알 수 있다.



<그림 7> 공정 관리한계선 변화

4. 결 론

본 연구는 통계적 공정관리 기법을 통하여 공정을 모니터링하고 공정의 이상이 발생할 경우 데이터마이닝 기법을 활용하여 공정변수 분류, 패턴파악, 예측을 통한 공정 개선을 제시하고 다시 공정으로 피드백 하여 공정을 모니터링 함으로서 지속적으로 공정을 관리할 수 있도록 하는 모형을 제시하였다. 본 연구에서 제시하는 분석 절차 모형을 통하여 실시간으로 수집되는 대용량을 공정데이터를 빠르게 분석하여 공정의 이상원인을 파악하고 공정의 안정을 유지, 관리하는 공정 개선방법을 제시함으로써 문제해결에 소요되는 시간과 비용을 줄일 수 있고 더 나아가서는 기업의 생산성을 향상시킬 수 있다.

## 참고문헌

- [1] 강형철 외 5명; “데이터마이닝”, 자유아카데미, 서울, 1999.
- [2] 변성규, 강창욱; “데이터마이닝 기법을 이용한 제조 공정내의 불량항목별 예측방법”, 한국산업 경영시스템학회지, Vol. 27, No. 2, pp. 10-16 2004.
- [3] 백동현, 한창희; “데이터 마이닝을 이용한 반도체 FAB공정의 수율개선 및 예측”, 한국지능정보시스템학회 논문지, Vol. 9, No. 1, pp. 157-177, 2003.
- [4] Braha, D., Shmilovici, A.; “Data mining for improving a cleaning process in the semiconductor industry”, IEEE transactions on semiconductor manufacturing, Vol. 15, No. 1, pp. 91-101, 2002.
- [5] Corinna, C., Vapnik, V.; “Support-Vector Network”, Machine Learning, Vol. 82, No. 14, pp. 273-297, 1995.
- [6] [http : svmlight.joachims.org](http://svmlight.joachims.org). Support Vector Machine, Computer Software, Collaborative Research Center, Thorsten Joachims, 2004.
- [7] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C.; “Learning algorithms for classification : comparison on handwritten digit recognition”, Theory neural networks : e statistical mechanics perspective, pp. 261-276, 1995.
- [8] Jiawei H., Micheline K.; “Data Mining : oncepts and Techniques”, Morgan kaufmann publishers, NY, 2001.
- [9] Montgomery, D. C.; “Introduction to Statistical Quality Control”, 4th ed., John Wiley & Sons, NY, 2001.
- [10] Timothy, J. S., Julie, K. S., Tomas, V.; “The application of artificial neural networks to monitoring and control of an induction hardening process”, Journal of Industrial Technology, Vol. 16, No. 1, pp. 1-11, 2000.
- [11] Vapnik, V.; “The Nature of Statical Learning Theory”, Springer, NY, 1995.
- [12] Zaki, M. J.; “Efficient enumeration of frequent sequences”, Conference on Information and Knowledge Management, Bethesda, Maryland, pp. 68-75, 1998.