

특성 변동 관리에 기반한 지능적 수율관리 방안

A new Intelligent Yield Management Methodology based on Feature Manipulation

이 장 희
한국기술교육대학교 산업경영학부

Abstract

This study presents a new intelligent yield management methodology which can forecast the yield level of a production unit based on features' behaviors. In this proposed methodology, we identify the existing features using C5.0 that are combination of nodes (i.e., variables) in the decision tree generated by C5.0, use SOM (Self-Organizing Map) neural networks in order to extract the feature's patterns and classify, and then make features' control rules using C5.0.

1. 서론

투입 수에 대한 양품의 비율을 의미하는 수율은 기업의 생산성, 수익성 및 업체의 성과 측면에서 매우 중요한 척도로 기업은 체계적인 수율관리를 위해 많은 노력을 기울이고 있다.

관리도, 실험계획법, 다변량 분석, 상관 및 회귀분석등과 같은 통계적인 기법을 적용하여 수율 및 품질 관리를 하거나 품질정보시스템을 구축하여 품질 및 수율변화를 사전에 예측하여 검출할 수 있는 시스템을 활용하기도 한다. 이와 같은 수율 및 품질관리 시스템에는 수많은 공정 변수와 대량의 데이터를 효과적으로 처리할 수 있는 데이터 마이닝 기법이 최근 제조 현장에 많이 활용되고 있기도 하다 [1]. 또한 반도체 제조 기업들처럼 업체의 조건에 맞게 적정한 수율 모델을 개발, 적용하여 수율예측을 통한 이상 수율을 예방하는 활동을 전개하기도 한다[2].

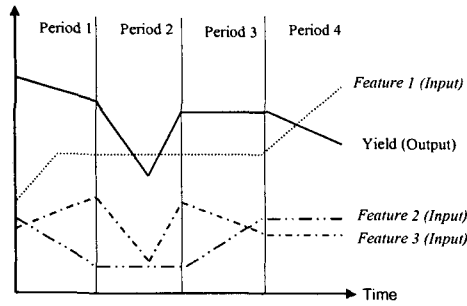
이러한 다방면의 활동에도 불구하고 수율이 워낙 수많은 변수들과 관련 있을 뿐 아니라 비선형적인 복잡한 상관 관계를 가지는 경우가 많기에 효과적인 수율관리 방안을 찾기

매우 어려운 상황으로 개별 기업마다 많은 시행착오를 거쳐 효과적인 방법을 구축한다. 이러한 시행착오는 수율에 영향을 주는 변수가 많을수록 또한 그 관계가 비선형적인 관계와 같이 복잡할수록 크다고 할 수 있겠다.

본 연구는 수율과 연관된 변수의 개수가 많고 비선형적인 복잡한 관계가 존재하는 상황에서 수율에 영향을 주는 3차 또는 4차와 같은 고차의 교호작용 변수들을 규명하여 이들의 조합을 특성(Feature)이라 명하고 특성들을 관리함으로써 복잡한 수율도 간단하게 관리할 수 있는 방법을 제시한다.

특성의 개념은 주성분 분석에서의 주성분, 인자 분석에서의 인자와 같이 다수의 변수들이 상호 작용하여 수율 변동에 영향을 주는 변수들의 조합인 특성을 고려하여 추출하는 것이다.

<그림 1>은 시간의 흐름에 따른 수율 변동을 개별 변수들의 변동으로 설명하지 않고 상호 교호작용을 가지는 변수들의 조합인 특성(Feature)들의 변동으로 표현한 것으로 본 연구의 기본 개념을 보여주고 있다.

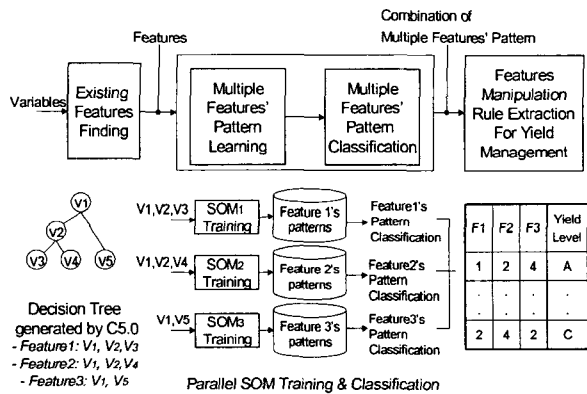


0) Feature 1 = Variable 1, Variable 2 및 Variable 7의 조합
 Feature 2 = Variable 1, Variable 3 및 Variable 5의 조합
 Feature 3 = Variable 2, Variable 4 및 Variable 9의 조합

<그림 1> 특성변동에 따른 수율변동 개념

2. 특성 기반 수율관리 방안

본 연구에서 제시하는 특성 관리에 기반한 수율 변동 관리 방안은 <그림 2>와 같다. 즉, 효과적인 수율변동 관리를 위해 수율변동에 영향을 주는 상호 교호작용을 가지는 변수들의 조합인 특성을 추출하고 이들의 변동 관리 규칙을 추출하여 특성관리를 하는 방법이다.



<그림 2> 수율변동 관리를 위한 특성관리 방안

본 연구에서 제안하는 방안의 1단계는 수율과 수율에 영향을 줄 수 있는 모든 변수로 구성된 데이터 셋의 분석을 통해 수율에 영향을 주는 특성들을 규명하는 것이다. 이를 위해, C5.0 분석을 수행하여 얻어지는 의사결정 나무의 노드 (즉, 변수) 구조에 기반하여 특성을 구성한다[3].

예를 들면, <그림 2>의 C5.0 분석을 수행하여 얻어진 의사결정 나무에는 5개의 노드 (V1,...,V5)가 있는데 노드간 구조에 기반하여 특성 1은 V1,V2, V3의 3가지 변수로 구성하고 특성2는 V1,V2,V4

의 3가지 변수로, 특성3은 V1,V5의 2가지 변수로 구성한다.

다음 2단계는 1단계에서 얻어진 각각의 특성들에 대해 기존 데이터에 존재하는 모든 패턴을 추출하는 것이다. 이를 위해 각 특성별로, 특성에 속하는 변수들로 구성된 데이터 셋을 구성하고 특성에 속하는 변수들의 움직임의 패턴을 SOM (Self-Organizing Map) 신경망을 이용하여 학습한다[4]. <그림 2>의 예에서, 특성 1에 속하는 3가지 변수인 (V1,V2, V3), 특성 2에 속하는 3가지 변수인 (V1,V2, V4), 특성 3에 속하는 2가지 변수인 (V1,V5)에 대해 각각 SOM 신경망을 구성하여 각 특성들의 패턴을 추출한다.

3단계는 기존 변수들의 데이터를 2단계에서 추출한 각 특성의 패턴을 기준으로 패턴 분류하여 특성 값으로 변환하는 것이다. 2단계에서 특성 1에 대해 9개의 패턴을 추출하였다고 가정하면, 기존 (V1,V2,V3)의 데이터를 특성1의 9개 패턴 중 어느 패턴에 속하는지 분류하는 것이다.

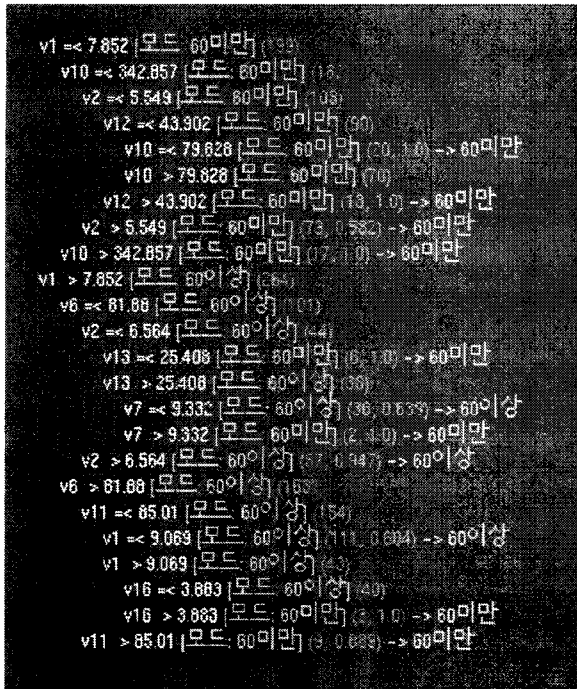
마지막 4단계는 3단계에서 분류한 각 특성의 패턴 값과 수율 변동간의 상관 관계를 분석하여 수율관리를 위한 특성의 변동 관리 규칙을 추출하는 것이다. <그림 2>에서 보인 것과 같이 3개의 특성별로 패턴 값과 그때의 수율 수준을 데이터 셋으로 구성하여 C5.0 분석을 수행하면 수율 변동에 대한 특성 관리 규칙을 얻을 수 있다.

3. 반도체 수율 관리에의 적용

2장에서 제시한 방법의 설명을 위해 반도체 수율 데이터 및 수율에 영향을 주는 변수 데이터를 확보하여 2장에서 제시한 방법에 의해 분석하였다. 현장 데이터 확보의 어려움 때문에 분석대상 데이터 셋은 수율에 영향을 주는 변수 15개 (v1,...,v15)와 수율 데이터가 463개 (Lot 단위) 레코드로 구성되었다.

2장에서 제시한 방안의 1단계 C5.0 분석을 위해 60% Lot 수율을 기준으로 60% 미만의 수율을 가지는 Lot을 '60미만', 60% 이상인 수율을 가지는 Lot을 '60이상'으로 표기하여 수율 수준을 구분하였다. 수율 60%는 당시 Lot의 이상 수율 판정을 하는 기준 값으로 현장에서 사용하고 있었다.

<그림 3>은 463개의 Lot 데이터에 대해 Clementine의 C5.0 분석을 수행하여 얻은 의사결정 나무이다.



<그림 3> 1단계 C5.0 분석결과

<그림 3>의 의사결정 나무 구조에 근거하여 다음과 같은 8개의 특성을 정의하였다.

- 1) 특성 1 (f1): V1, V10, V2, V12의 조합
- 2) 특성 2 (f2): V1, V10, V2의 조합
- 3) 특성 3 (f3): V1, V10의 조합
- 4) 특성 4 (f4): V1, V6, V2, V13의 조합
- 5) 특성 5 (f5): V1, V6, V2, V13, V7의 조합
- 6) 특성 6 (f6): V1, V6, V2의 조합
- 7) 특성 7 (f7): V1, V6, V11의 조합
- 8) 특성 8 (f8): V1, V6, V11, V16의 조합.

8개 각 특성별로 패턴 추출을 위해 8개의 SOM 신경망 모델을 구축하였다. 즉, 특성 1의 경우 453개의 Lot에 존재하는 (V1, V10, V2, V12)의 4가지 변수들의 동시적인 움직임의 패턴을 추출하고자 SOM 신경망 모델을 구축하는 것이다.

이때, 4가지 변수의 동시적인 움직임의 패턴을 9가지 패턴 ('0', '1', '2', '10', '11', '12', '20', '21', '22')으로 구분하여 추출하였다 (<표 1>참조).

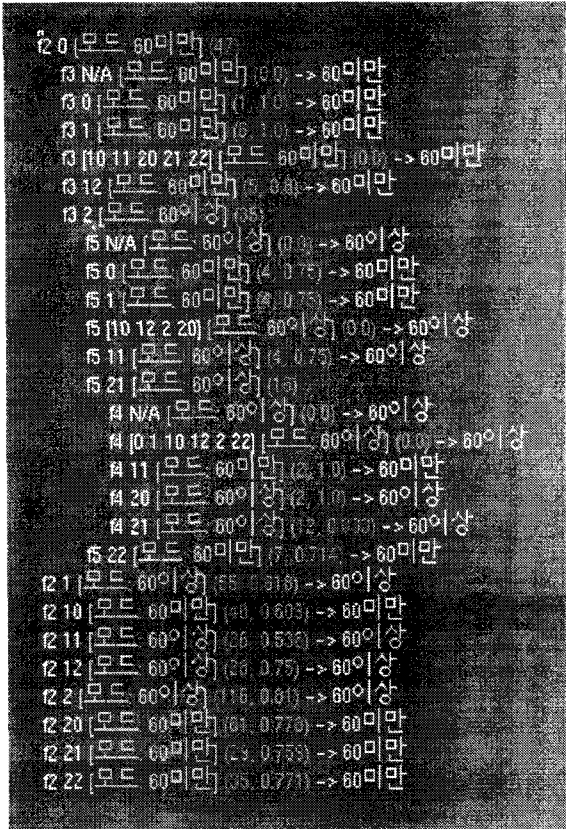
9가지 패턴	해당 패턴에 속하는 Lot수
0	99
1	52
2	51
10	29
11	22
12	44
20	65
21	28
22	73
Total	463

<표 1> SOM 신경망에 의한 특성1의 패턴 분석 요약

다음 단계는 463개 Lot에 대해 각 Lot이 어느 패턴에 속하는지를 규명하는 것인데, 예를 들면 (V1, V10, V2, V12) = (9.71, 6.81, 91.90, 13.42)의 값을 가지는 Lot1은 패턴 '0'에 해당하고 패턴 '0'에 해당되는 Lot이 총 99개 있음을 알 수 있다. 특성2를 비롯한 다른 7개의 특성들에 대해, 특성1에 대해 수행한 이상의 작업을 동일하게 실시한다.

이상의 작업을 완료하면 (V1, ..., V15)의 값으로 표현된 463개 Lot (특성1, 특성2, ..., 특성8)의 값으로 변환하여 표현할 수 있다. 이때, 특성1을 비롯한 8개의 특성 값은 모두 '0', '1', '2', '10', '11', '12', '20', '21', '22'의 9개 값 중 하나로 표현되지만 특성별로 값이 같더라도 서로 다른 의미를 가진다. 즉 특성1과 특성2의 값이 '0'으로 동일하다 할지라도 두 특성의 패턴이 같은 것은 아니다. 이상에서 설명한 방식으로 데이터를 정리하면, Lot1은 (특성1, 특성2, ..., 특성7, 특성8) = (0, 2, 0, 0, 2, 2, 2, 2, 1)로 표현된다.

마지막 단계로, 특성 값으로 표현된 463개 Lot 데이터와 그때의 수율 수준 (60이상인 Lot과 60미만인 Lot) 데이터에 대해 C5.0 분석을 수행하여 다음 <그림 4>와 같은 의사결정 나무를 얻었다.



<그림 4> 4단계 C5.0 분석결과

<그림 4>의 의사결정나무로부터, 다음과 같은 신뢰도가 0.8 이상인 6개의 특성 관리 규칙(RULE)을 추출하였다.

- 1) 규칙1(신뢰도 100%): (특성2 =0) 이고 (특성3 =0) 이면 수율이 '60미만'
- 2) 규칙2(신뢰도 100%): (특성2 =0) 이고 (특성3 =1) 이면 수율이 '60미만'
- 3) 규칙3(신뢰도 80%): (특성2 =0) 이고 (특성3 =12) 이면 수율이 '60미만'
- 4) 규칙4(신뢰도 100%): (특성2 =0) 이고 (특성3 =2) 이고 (특성5 =21)이고 (특성4 =11) 이면 수율이 '60미만'
- 5) 규칙5(신뢰도 100%): (특성2 =0) 이고 (특성3 =2) 이고 (특성5 =21)이고 (특성4 =20) 이면 수율이 '60이상'
- 6) 규칙6(신뢰도 81%): (특성2 =2) 이면 수율이 '60이상'

상기한 6가지 규칙을 보면, 특성2의 패턴이 '0'일 때 수율이 대부분 60미만인 경우가 많음을 알 수 있다. 특성2의 패턴이 '0'인 경우는 v1평균이 8.1445로 9개 패턴 중에서 중간 정도, v2평균이 6.0974로 9개 패턴 중에서

낮은 정도, v10평균이 106.72로 9개 패턴 중에서 낮은 정도에 해당한다 (<표 2> 참조).

그룹	v1평균	v2평균	v10평균
0	8.1445	6.0974	106.72
1	8.6302	6.7192	106.21
2	9.0615	7.1079	102.38
10	7.0138	5.347	133.03
11	7.8304	6.0447	105.03
12	8.4889	6.7757	109.72
20	6.596	4.8411	298.87
21	7.165	5.315	179.98
22	7.9416	6.1128	123.48

<표 2> 특성2의 9개 그룹별 평균값

실제 현장에서 3가지 변수가 상기와 같은 특징을 가질 때, 수율이 낮아지는 현상을 확인할 바 있다.

4. 결과 및 향후 연구

본 연구는 수율에 영향을 주는 변수들이 많이 존재하고 그들 간의 복잡한 상호 작용에 의한 수율변동이 예상되는 환경에서 상호 작용을 가지는 변수들의 조합을 특성으로 정의하고 이들의 변동을 관리함으로써 수율변동을 효과적이고 쉽게 관리할 수 있는 방안을 제시하였다. 특성 구성 및 특성변동의 관리 규칙 추출을 위해 C5.0 분석을 수행하였고, 특성별 패턴 추출 및 패턴 규명을 위해 SOM 신경망을 활용하였다. 향후 본 논문에서 제시한 방법의 타당성 검증 등을 위해 좀 더 많은 현장 데이터를 확보하여 적용하고자 한다.

참고문헌

- [1] 백동현, 남정곤, 반도체 일관가공 공정의 수율개선 및 예측을 위한 데이터마이닝 기법과 OLAP의 활용, 한국경영정보학회 춘계학술대회, 2002
- [2] 김수영, 박광수, 전치혁, 반도체 제조업에서 사용되는 수율 모델의 비교 및 이용, 대한산업공학회, 1997
- [3] Quinlan, J. Ross, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers(San Mateo, California), 1993
- [4] Kohonen, T. ,*Self-Organization & Associative Memory*. 3rd ed. (Springer-Verlag, Berlin), 1989.

