

# 반도체 공정의 이상 탐지와 분류를 위한 특징 기반 의사결정 트리

손지훈 · 고종명 · 김창욱<sup>\*</sup>

연세대학교 정보산업공학과

## Feature Based Decision Tree Model for Fault Detection and Classification of Semiconductor Process

Ji Hun Son · Jong Myoung Ko · Chang Ouk Kim

Department of Information and Industrial Engineering, Yonsei University, Seoul, 120-749

As product quality and yield are essential factors in semiconductor manufacturing, monitoring the main manufacturing steps is a critical task. For the purpose, FDC(Fault detection and classification) is used for diagnosing fault states in the processes by monitoring data stream collected by equipment sensors. This paper proposes an FDC model based on decision tree which provides if-then classification rules for causal analysis of the processing results. Unlike previous decision tree approaches, we reflect the structural aspect of the data stream to FDC. For this, we segment the data stream into multiple subregions, define structural features for each subregion, and select the features which have high relevance to results of the process and low redundancy to other features. As the result, we can construct simple, but highly accurate FDC model. Experiments using the data stream collected from etching process show that the proposed method is able to classify normal/abnormal states with high accuracy.

**Keyword:** semiconductor manufacturing, fault detection and classification, decision tree, wrapper based feature selection

### 1. 서 론

#### 1.1 연구 배경

최근 주요 반도체 제조 공정은 300mm 웨이퍼(Wafer) 제조 설비의 도입이 주류를 이룸에 따라서 높은 생산성을 달성하였다. 하지만 고집적 반도체 제품을 생산하기 위한 정교한 제조 설비들의 안정적인 운영은 그만큼 복잡해 졌으며, 공정 중에 폐기 처분되는 불량 발생 위험도 증가하였다. 따라서 반도체 제조 공정에서 웨이퍼의 불량 발생을 최소화하고 생산량을 증대시키기 위해서는 민감하게 공정의 이상을 탐지하고 이상의 원

인을 분류(FDC : Fault detection and classification)하여 이에 대한 신속한 조치를 취하는 것이 무엇보다 중요하다(Goodlin *et al.*, 2003).

반도체 제조 공정들은 서로 유기적으로 연결되어있기 때문에 특정 공정에서 발생한 이상은 연결된 공정들에 직접적으로 영향을 미친다. 또한 반도체 제품이 완성되기까지는 긴 리드타임이 소요되기 때문에 특정 공정에서 발생한 이상을 간과할 경우 많은 시간 및 비용 손실이 발생한다. 따라서 대부분의 공정들이 센서들을 부착하여 모니터링 되고 있다. 이러한 센서들은 전력, 가스 유입량, 온도, 그리고 압력 등과 같은 공정 변수들의 상태를 일정 시간 간격으로 측정한다. 웨이퍼 단위로

<sup>\*</sup>연락처 : 김창욱 교수, 120-749 서울시 서대문구 신촌동 134 연세대학교 정보산업공학과, Fax : 02-364-7807, E-mail : kimco@yonsei.ac.kr  
투고일(2008년 12월 10일), 심사일(1차 : 2009년 01월 22일, 2차 : 2009년 02월 11일), 게재확정일(2009년 02월 25일).

얻어진 센서 데이터들로 이루어진 데이터 스트림(Data stream)은 각 제조 조건(Recipe)에 따라 특정한 패턴을 보여준다. 정상 상태에서 얻어진 데이터 스트림들이 보여주는 패턴은 유사한 구조적 형태를 갖는다.

현재 이상 탐지를 위한 통계적 분석 기법들은 데이터 스트림을 구성하는 전체 센서 신호의 구조적인 특성을 반영하기 보다는 특정 시점들에서 추출된 값들의 집합을 공정의 대표적인 상태로 간주한다. 그리고 이러한 값들을 사용하여 공정 변수들의 평균, 표준편차, 그리고 주성분 분석(Principle component analysis), Hotelling의  $T^2$  통계치와 같은 다변량 모델의 매개변수 설정을 위한 공분산과 같은 요약된 통계치를 유도한다(Dunia *et al.*, 1996; Cherry and Qin, 2006; Guo *et al.*, 1991). 특히 이러한 다변량 모델들은 센서 측정치들이 다변량 정규 분포를 따를 경우 좋은 성능을 보이는 것으로 알려져 있다. 한편, 레퍼런스 패턴 기반의 이상 탐지 수행 연구는 공정이 정상 상태일 경우 데이터 스트림의 형태가 유사한 모양을 갖는다는 것을 근거로 한다. Salvador *et al.*(2004)는 공정의 각 단계별로 지켜야 할 조건을 상태(State)로 정의하고 입력되는 데이터 스트림을 통해 비정상 상태로 전이되는 여부를 판단하였다. Lada *et al.*(2002)는 정상상태의 데이터 스트림을 웨이블릿(Wavelet) 계수를 이용하여 표현하였다.

## 1.2 연구 동기 및 목적

의사결정 트리(Decision tree)는 “if-then” 분류 규칙들의 집합으로 표현될 수 있기 때문에 해석과 인과관계 분석이 용이하다. 가령, if-절들은 설비 센서로부터 추출된 공정 변수들의 측정값들이 되고 then-절들은 해당되는 공정의 결과가 된다. 즉, 새로운 공정이 진행되면서 관측된 공정의 상태는 구축된 의사결정 트리에 의해 공정 상태에 대한 결과에 맵핑된다. 이런 이유에서 의사결정 트리는 신경망(Neural networks),  $k$ -인접 이웃 분류 규칙( $k$ -Nearest neighbor rule)과 같은 기계 학습(Machine learning) 기법 모델들과는 달리 학습 결과를 쉽게 이해할 수 있는 형태로 제공한다.

의사결정 트리 학습은 지도학습을 통한 귀납적 분류에 의해 이루어지며 트리 구축에 필요한 데이터가 다변량 정규 분포로부터 추출되는 것을 필요로 하지 않는다. 의사결정 트리를 FDC에 적용한 초기 연구인 Irani *et al.*(1993) 그리고 Baek *et al.*(2002)은 각 설비 센서들에서 얻어진 공정의 대표적인 상태값들과 오프라인 계측장비(Metrology)에 의해 측정된 제품의 실제 상태 정보(목표값)를 이용하여 원인-결과 지식을 학습하였다. 오프라인 방식에 의해 제품의 목표값을 획득하기 위해서는 많은 비용과 시간이 소모됨에도 불구하고, 학습된 의사결정 트리는 공정의 어떤 부분이 각각의 제품 결함에 대한 주원인으로 작용했는지 직관적으로 나타낼 수 있는 큰 장점이 있다.

본 논문에서는 기존의 의사결정 트리 기반의 연구들과는 달리, 반도체 제조 공정의 FDC를 위해 웨이퍼 단위 데이터 스트

림이 나타내는 구조적인 측면을 반영하였다. 이를 위해, 공정의 결과에 높은 관련성을 가지고 있는 데이터 스트림의 구조적인 특징들을 추출하고 의사결정 트리를 구축하는데 이들을 입력 요소로 사용한다. 본 논문에서 특징들은 공정 변수들의 평균과 표준편차, 공정의 지속 시간, 그리고 공정 변수들의 이상치(최대값 그리고 최소값)로 5 가지의 형태로 구분한다. 또한 민감한 이상 탐지를 위해, 데이터 스트림의 변화 시점들을 중심으로 복수개의 세부 영역으로 분할하고 각 세부 영역에 대해 동일한 형태의 특징들을 정의하였다. 의사결정 트리의 구조에서, 각각의 내부 노드는 이상 진단을 위해 선택된 특징을 나타내며, 특징의 경계점(Threshold)에 따라 자식 노드로 분할된다. 최하위 노드는 공정의 결과(정상 또는 결함)를 나타낸다. 최상위 노드로부터 최하위 노드까지의 경로는 공정의 결과를 나타내는 특징들의 조합이며 이러한 경로가 바로 분류 규칙이 된다.

많은 수의 특징들은 의사결정 트리의 크기를 불필요하게 증가시켜 공정에 대한 분석을 매우 복잡하게 만들 수 있다. 예들 들어, 4개의 세부 영역으로 구성된 단일 센서의 신호만 고려하여도 전체 특징의 수가 25개가 되며, 사용되는 공정 변수의 증가는 더 많은 특징수를 만들어낸다. 따라서 간결하면서도 정확한 분류 성능을 지닌 의사결정 트리를 만들기 위해, 본 논문에서는 특징 선택(Feature selection) 과정을 통해, 공정의 결과에 대해 예측력이 없거나 작고 그리고 특징들 간의 중복성을 지닌 요소들을 제거하였다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 특징 기반 의사결정 트리를 이용한 FDC 모델을 소개하며 제 3장에서는 제안한 모델의 성능을 실험을 통해 검증한다. 제 4장에서는 결론을 내린다.

## 2. 특징 기반 의사결정 트리

### 2.1 구조적 특징 추출을 위한 데이터 스트림의 분할 및 특징 정의

<그림 1(a)>는 주요 반도체 제조 공정 중 하나인 에칭 공정에서 얻어진 전압 센서에서 추출된 데이터 스트림이다. 데이터 스트림은 복수개의 영역으로 분할될 수 있는 구조적 형태를 보인다. 공정이 시작된 후 일정시간 비공정 상태(Idle state)가 유지된다는 가정 하(Choi *et al.*, 2007)에 데이터 스트림을 구성하는 신호가 비공정 상태로 전이되는 시점을 변화시점으로 인식하여 영역을 구분한다. 예시된 데이터 스트림의 경우에는 정의한 기준에 의거하여 <그림 2(b)>에서처럼 7개의 세부 영역으로 분할된다.

데이터 스트림으로부터 추출되는 특징들은 <표 1>에서 볼 수 있듯이 광역 특징(Global feature)과 지역 특징(Local feature)으로 정의한다. 광역 특징은 한 웨이퍼 단위의 데이터 스트림의

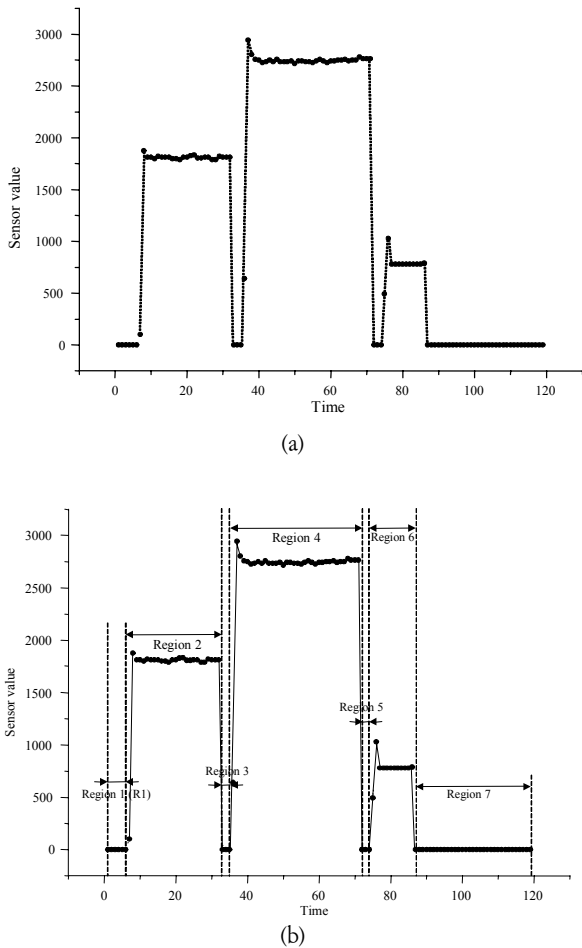


Figure 1. 전압 센서의 데이터 스트림

신호를 기준으로 그리고 지역 특징은 데이터 스트림을 분할한 각 세부 영역별로 정의된다. <그림 1>의 데이터 스트림의 경우 7개의 영역으로 구분되므로 총 40개의 특징들이 정의된다.

Table 1. 구조적 특징의 정의

유형	특징
광역특징	전체 길이, 평균 값과 표준편차, 최대 값과 최소 값
지역특징	각 영역별 길이, 각 영역내 신호들의 평균 값과 표준편차 및 최대 값과 최소 값

### 2.2 의사결정 트리의 구축

의사결정 트리의 학습과정은 훈련 데이터 집합(Training dataset)을 이용하여 분류 에러가 최소화 될 수 있도록 특징들의 공간을 재귀적으로 분할하는 것이다. 학습된 의사결정 트리는 각각 해체될 수 있는 하이퍼큐브(Hyper cube)의 집합과 같다. 분류 에러는 각 큐브들의 대부분이 동일한 클래스를 가질 때 최소화된다. 분류 에러를 최소화하기 위해, 탑-다운(Top-down) 그리고 그리디 알고리즘(Greedy algorithm)을 이용한다. 분류 에러

를 평가하기 위해 엔트로피(Entropy) 개념이 사용된다. 엔트로피는 집합의 불순수성 정도를 의미하고 식 (1)에 의해 측정된다. 단,  $k$ 는 클래스의 수,  $p_i$ 는 훈련 데이터 집합  $S$ 에서 클래스  $i$ 의 비율이다.

$$Entropy(S) = - \sum_{i=1}^k p_i \log p_i \quad (1)$$

최상위 노드로부터 하위 방향으로 검증 속성(test attribute)을 선택하게 되는데 식 (2)의 정보 획득량(Information gain)을 측정하여 가장 큰 정보 획득을 가져오는 속성부터 선택된다. 즉, 전체 트리의 불순수성을 줄이는 방향으로 의사결정 트리가 형성된다. 최상위 노드로부터 모든 하부 노드에 이르기까지 탐-다운 방식으로 수행된다.  $value(A)$ 는 속성  $A$ 에 대한 모든 가능한 값들의 집합이며,  $S_v = \{s \in S | A(s) = v\}$ 이다.  $Gain(S, A)$ 는 훈련 데이터 집합  $S$ 를 속성  $A$ 의 값에 의해 분할할 때, 얻어지는 정보의 획득량이다.  $|S|$ 는  $S$ 에 속하는 개별 데이터의 개수이다.

$$Gain(S, A) = Entropy(S) - \sum_{v \in value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

일부 특징들은 연속형 값을 갖는다. 의사결정 트리는 기본적으로 이산형 값을 갖는 특징들을 사용하는 학습방법이다 (Quinlan, 1986). 따라서 연속형 값을 갖는 특징들은 각 특징들이 갖는 값의 범위를 다수의 구간으로 분할하여 이산화되어야 한다. 이러한 다구간 이산화 과정은 사용된 훈련 데이터 집합에만 한정적으로 과적합(Overfitting)되는 의사결정 트리를 형성할 수 있다. 즉, 알려지지 않은 데이터들에 대한 분류 에러를 가질 수 있다는 의미이다. 의사결정 트리가 과최적화되는 것을 방지하기 위해, 본 논문에서는 MDL(Minimum description length) 원리(Principle)에 기반한 다구간 이산화 과정을 수행한다 (Fayyad and Irani, 1993).

### 2.3 구조적 특징의 선택

특징 수가 많으면, FDC 모델이 복잡해지므로 이상 탐지와 분류에 영향을 미치는 중요한 특징들만 고려하여 의사결정 트리를 구축해야 한다. 특징 선택은 목표값과 관련성(Relevance)이 떨어지거나 서로 중복성(Redundancy)을 갖는 특징들을 제거하여 주요한 소수의 특징들을 자동으로 발견하는데 유용하다 (Guyon and Elisseeff, 2003). 즉, 목표값과 상호 의존성이 높은 특징들을 선택해야 하며, 선택된 특징들 간에는 낮은 상호 의존성을 가져야 한다. 상호 의존성은 식 (3)의 상호정보(Mutual information)에 의해 측정한다. 만약 두 개의 이산형 확률 변수  $X$ ,  $Y$ 가 있을 경우 상호 정보  $I(X; Y)$ 는 식 (3)과 같이 계산된다.

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \quad (3)$$

여기서  $p(x, y)$ 는  $X, Y$ 의 결합 확률 분포 함수(Joint probability distribution function)이고  $p(x), p(y)$ 는  $X$  그리고  $Y$ 의 주변 확률 분포 함수(Marginal probability distribution function)이다. 참고로 본 연구에서 사용된 특징들은 특징들 간의 중복성은 0.247에서 0.021정도의 범위 그리고 목표값과 특징들의 관련성은 0.036에서 0.289정도의 범위를 가졌다. 본 연구에서 목표값은 공정의 정상과 비정상 두 가지 클래스로 이루어진다. <그림 2>에서  $Avg_T, Std_T, Dur_T$  그리고  $Max_{R6}$ 는 각각 데이터 스트림을 구성하는 전체 신호들의 평균, 표준편차, 길이 그리고 6번째 세부 영역에 속한 신호들 중 최대값이다.

<그림 2(a)>는 본 연구에서 사용된 특징들 간의 중복성을 나타낸다.  $Avg_T$ 와  $Std_T$ 의 경우 상대적으로 높은 상호 의존성 즉, 높은 중복성(0.247)을 가지며  $Dur_T$ 와  $Max_{R6}$ 는 매우 낮은 상호 의존성을 가지고 있어 중복성(0.080)이 낮음을 알 수 있다. <그림 2(b)>는 목표값의 각 클래스와 사용된 특징 간의 상호 의존성 즉, 관련성을 나타낸다.  $Dur_T$ 는 목표값의 클래스에 거의 1:1로 매칭되어 관련성이 높게 나온 경우(0.289)이며  $Max_{R6}$ 는 여러 부분에서 중복되어 관련성이 매우 낮은 관련성(0.036)을 갖는다.

FDC 모델을 간결하면서 정확성 높게 구축하기 위해 본 논문에서는 낮은 복잡도와 뛰어난 분류 성능을 보장하는 랩퍼

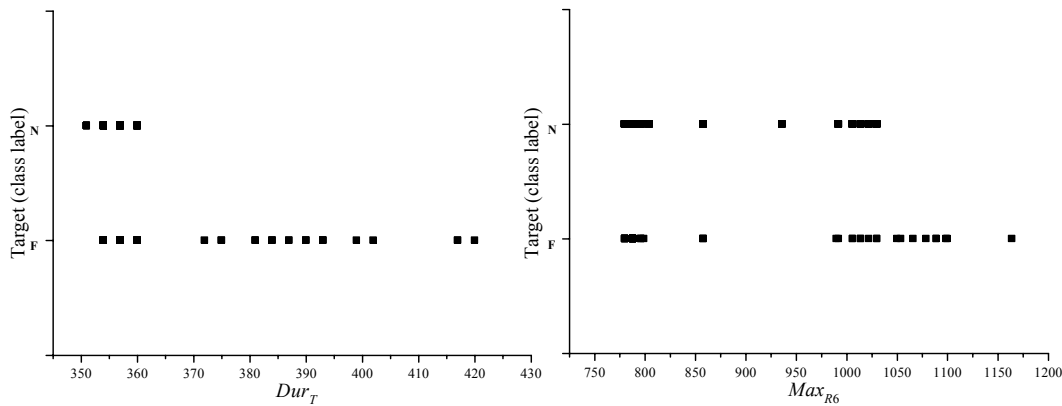
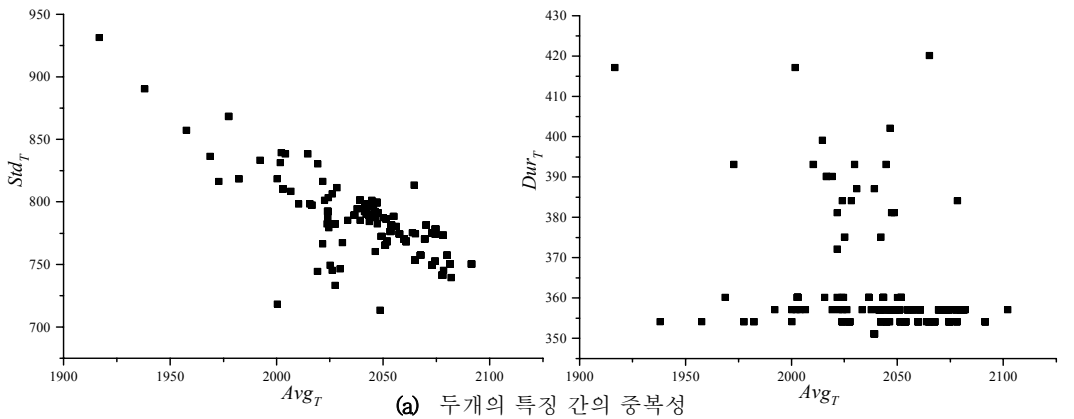
(Wrapper)를 이용한 특징 선택 방법(Kohavi and John, 1997)을 적용하여 주요한 특징들을 선택한다. 랩퍼는 직접적으로 의사결정 트리의 분류 정확도를 측정해가면서 최적의 특징들을 탐색하는 방법이다. 랩퍼는 전방 선택과 후방 선택의 두 가지 탐색 구조를 가지고 있다.

#### (1) 전방 선택(Forward search)

총  $k$ 개의 특징이 있을 경우 전방 선택에 의한 특징 선택 알고리즘은 다음과 같다. 전방 선택은 한 번에 한 개씩 특징을 추가해나가는 방식이다.  $f_j$ 는  $j$ 번째로 선택된 특징을 의미한다.

**단계 1:** 추출된 전체 특징 집합으로부터 한 개의 특징으로 구성된  $k$ 개의 의사결정 트리를 생성하고 이들 중 분류 정확도가 가장 높은 특징을 첫 번째 특징,  $f_1$ 으로 선택한다.

**단계 2:**  $j-1$ 개의 특징들( $f_1, f_2, \dots, f_{j-1}$ )이 이미 선택된 경우,  $j$ 번째 특징은 남아있는 특징들을 이용하여  $k-j+1$ 개의 의사결정 트리를 생성한 후 이들 중 분류 정확도가 가장 높은 것을 생성하는 데 사용된 특징을  $f_j$ 로 선택한다. 즉, <그림 3>에서 보듯이  $r_i$  ( $i = j, j+1, \dots, k$ )가 선택되지 않은 특징들일 경우 각각  $j$ 개의 특징들로 이루어진  $k-j+1$ 개의 의사결정



(b) 한 특징과 목표값 간의 관련성

Figure 2. 중복성과 관련성의 측정

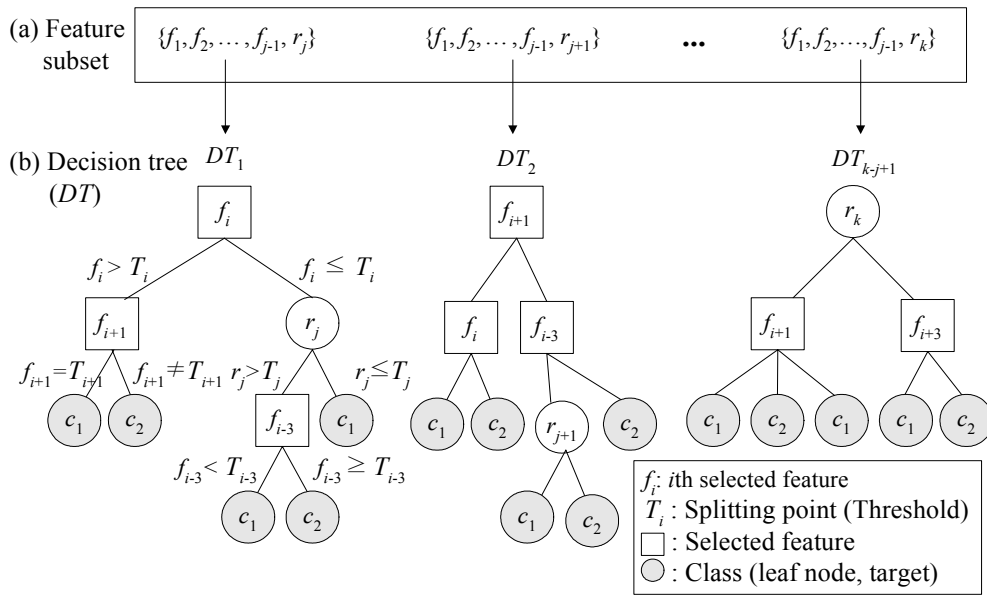


Figure 3. 전방 탐색을 통한 특징 선택

정 트리 중에서 분류 정확도가 가장 높은 것을 선택한다.

**단계 3:** 만약 분류 정확도가 개선되었다면 단계 2로 이동하여 추가적으로 특징들을 선택하고, 그렇지 않은 경우에는 전 시행까지 추출된 특징들의 부분 집합을 이용하여 의사결정 트리를 구축한다(현 시행에서 추가된 특징은 포함시키지 않음).

(2) 후방 선택(Backward search)

후방 선택은 최초 모든 특징들로 이루어진 집합에서 한 번에 한 개씩 특징을 제거해나가는 방식이다.

**단계 1:** 전체 특징 집합으로부터 한 개의 특징을 제거한 부분 집합으로 구성된  $k$ 개의 의사결정 트리를 생성하고 이들 중 분류 정확도가 가장 높은 트리를 구축할 때 제거된 특징을 첫 번째로 제거한다.

**단계 2:**  $j-1$ 개의 특징들이 이미 제거된 경우  $j$ 번째 제거되는 특징은 남아있는  $k-j+1$ 개의 특징들을 각각 하나씩 제거하여 만든  $k-j+1$ 개의 의사결정 트리들 중 분류 정확도가 가장 높은 트리를 구축할 때 제거된 특징을 제거한다.

**단계 3:** 만약 분류 정확도가 개선되었다면 단계 2로 이동하여 추가적으로 특징들을 제거하고, 그렇지 않은 경우에는 전 시행까지 남아있는 특징들의 부분 집합을 이용하여 의사결정 트리를 구축한다.

새로운 특징들이 추가되면 기존 선택된 특징들이 설명하지 못하는 부분에 대한 설명이 가능하지만 반대로 기존 특징들이 설명하는 부분과 반대가 되는 경우가 발생할 수 있다. 이러한 사실은 전방 선택을 중지하고 최적의 특징 부분 집합을 확정

하는 시점을 결정해준다. 후방 선택에서도 특징의 제거가 반드시 설명할 부분의 일부를 간과하는 경우를 발생시켜 최적의 특징 부분 집합 생성의 기준을 결정해준다. 본 논문에서는 정확도 높고 간결한 공정 이상 진단 및 분류 규칙을 유도해내기 위해 전방 그리고 후방 선택 과정으로 선택된 특징들의 부분 집합들 중 분류 에러율이 가장 낮고 작은 특징들로 구성된 특징들의 부분 집합을 선택한다.

3. 실험

본 논문에서 제안된 FDC 모델을 검증하기 위해 주요 반도체 제조 공정인 에칭 공정에 부착된 센서들로부터 얻어진 데이터 스트림을 이용하였다. 해당 공정은 20여 개의 센서들을 통해 실시간으로 모니터링 되고 있었으며 이들은 전압, 전력, 기체 유입량, 온도 등 기능적 측면에서 분류될 수 있다. 센서의 패턴으로부터 다양한 구조적 특징을 정의하기 위해서 센서의 신호값이 공정의 세부 단계별로 구분되는 특성을 가지는 것은 필수적이다. 따라서 본 연구에서는 이러한 특성을 가지는 센서들 중 구조적 특징이 뚜렷한 전압 센서를 이용하여 이것으로부터 추출된 데이터 스트림을 실험에 사용하였다. 해당 센서는 <그림 1>과 같은 구조적 패턴을 가지고 있으며 이것으로부터 40개의 특징을 정의하였다. 데이터의 클래스는 정상과 이상 두 가지로 구분된다. 그리고 모델 구축을 위해 정상 310개, 이상 60개, 그리고 잡음(noise) 3개로 구성된 훈련 데이터 집합 그리고 구축된 모델의 테스트를 위한 정상 100 그리고 이상 20개의 데이터 스트림을 이용하였다.

실험은 크게 두 가지 측면에서 수행된다. 공정의 이상 진단을 위해 정확도 높은 분류 성능을 제공하는 FDC 모델 구축이

첫 번째이고 다음으로 특징 선택 과정을 통해 생성된 모델과 단순히 정의된 모든 특징들을 사용하여 구축된 모델을 비교하여 특징 선택 기반의 모델 구축의 중요성을 검증하는 것이다.

### 3.1 특징 선택 결과

랩퍼에 의한 특징 선택 결과는 <표 2>에서 보듯이 훈련 데이터 집합을 기준으로 한 낮은 분류 에러율을 가지면서도 간결한 특징의 부분 집합을 도출하였다. 전방 탐색의 경우 12개의 특징들이 선택되었으며, 후방 선택의 경우 7개의 특징들이 선택되었고, 분류 에러 측면에서도 후방 탐색이 더 우수한 결과를 나타냈다.

선택된 특징을 살펴보면 분할된 특징 영역 혹은 전체 영역의 표준편차, 평균, 길이, 최대값 등이 선택되었음을 알 수 있다. 선택된 특징들이 속하는 데이터 스트림 내의 영역을 살펴보면, 공정 조건에 따라 전압을 올려줘야 되는 부분들이 중요함을 알 수 있다. 또한 전방 선택의 경우 평균과 표준편차에 대한 특징들이 많이 선택되었는데 이는 전압을 올리고 유지하는 부분에서 안정적인 공정 관리가 필요함을 나타낸다. 후방 선택의 경우에는 전방 선택에 비하여 작은 수의 특징들과 더 낮은 분류 에러율을 나타냈지만 전체 40개의 특징에서 7개의 특징을 선택하기까지 더 많은 시행을 반복해야 하는 단점이 있다. 단적으로 전방 선택은 13회의 특징 선택 시행, 그리고 후방 선택은 34회의 특징 선택 시행을 거쳤다. 앞서 설명했듯이 최종적인 의사결정 트리는 후방 선택에 의해 선택된 7개의 특징들을 이용하여 작성하였다. 작성된 의사결정 트리는 테스트 집합에 대해 0.83%의 낮은 분류 에러를 나타냈다. 전방 선택에 의해 작성된 의사결정 트리는 테스트 집합에 대해 1.63%의 낮

Table 2. 랩퍼에 의한 특징 선택 결과

선택 순서	전방 선택	후방 선택
1	$Avg_T$	$Std_T$
2	$Max_{R4}$	$Dur_T$
3	$Std_T$	$Max_{R4}$
4	$Dur_T$	$Min_{R4}$
5	$Avg_{R2}$	$Avg_{R2}$
6	$Avg_{R4}$	$Max_{R6}$
7	$Max_{R2}$	$Std_{R6}$
8	$Std_{R6}$	
9	$Std_{R4}$	
10	$Avg_{R6}$	
11	$Dur_{R4}$	
12	$Std_{R2}$	
특징 수	12	7
분류 에러율	0.81%	0%

은 분류에러를 나타냈지만 상대적으로 복잡한 형태의 트리가 생성되었다.

### 3.2 특징 선택 여부에 따른 FDC 모델의 성능 비교

특징 선택 과정이 FDC 모델 성능에 어떠한 영향을 미치는지 확인하기 위해 특징 선택 과정을 통해 선택된 작은 수의 특징들을 사용하여 구축된 의사결정 트리와 정의된 모든 특징들을 사용하여 구축된 것과의 차이를 비교하였다. 특징 선택 과정 없이 의사결정 트리를 구축했을 경우에는 총 10개의 특징들

Table 3. 특징 선택 적용 시의 분류 규칙

규칙	내용	노드 수
1	IF $Std_T < 740.0$ , THEN class is abnormal.	1
2	IF $Std_T > 810.5$ , THEN class is abnormal.	1
3	IF $740.0 \leq Std_T \leq 810.5$ and $Dur_T > 366.0$ , THEN class is abnormal.	2
4	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ and $Max_{R4} < 2819.0$ , THEN class is abnormal.	3
5	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ and $Max_{R4} > 3006.0$ THEN class is abnormal.	3
6	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ and $Min_{R4} < 42.0$ THEN class is abnormal.	4
7	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $Min_{R4} \geq 42.0$ and $Avg_{R2} < 1693.9$ THEN class is abnormal.	5
8	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $Min_{R4} \geq 42.0$ , $Avg_{R2} \geq 1693.9$ and $Max_{R6} > 1040.0$ THEN class is abnormal.	6
9	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $Min_{R4} \geq 42.0$ , $Avg_{R2} \geq 1693.9$ , $Max_{R6} > 1040.0$ and $Std_{R6} < 192.6$ THEN class is abnormal.	7
10	IF $740.0 \leq Std_T \leq 810.5$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $Min_{R4} \geq 42.0$ , $Avg_{R2} \geq 1693.9$ , $Max_{R6} > 1040.0$ and $Std_{R6} \geq 192.6$ THEN class is normal.	7
규칙 수	10개	

Table 4. 특징 선택 적용 안했을 시의 분류 규칙

규칙	내 용	노드 수
1	IF $Avg_T > 2002.0$ , THEN class is abnormal.	1
2	IF $2002.0 \leq Avg_T < 2024.0$ , THEN class is normal.	1
3	IF $2024.0 \leq Avg_T < 2027.0$ and $Dur_{R2} > 84.0$ THEN class is abnormal.	2
4	IF $2024.0 \leq Avg_T < 2027.0$ , $Dur_{R2} \leq 84.0$ and $Avg_{R2} < 1693.0$ THEN class is abnormal.	3
5	IF $2024.0 \leq Avg_T < 2027.0$ , $Dur_{R2} \leq 84.0$ , $Avg_{R2} < 1693.0$ and $Avg_{R4} < 2026.0$ THEN class is abnormal.	4
6	IF $2024.0 \leq Avg_T < 2027.0$ , $Dur_{R2} \leq 84.0$ , $Avg_{R2} < 1693.0$ and $Avg_{R4} \geq 2026.0$ THEN class is normal.	4
7	IF $2027.0 \leq Avg_T < 2035.0$ , THEN class is abnormal.	1
8	IF $Avg_T \geq 2035.0$ and $Dur_T > 366.0$ , THEN class is abnormal.	2
9	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ and $Max_{R4} > 3006.0$ THEN class is abnormal.	3
10	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ and $Max_{R4} < 2819.0$ THEN class is abnormal.	3
11	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , and $Avg_{R6} < 716.0$ THEN class is abnormal.	4
12	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , and $Avg_{R6} > 804.0$ THEN class is abnormal.	4
13	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ and $Std_T > 810.0$ THEN class is abnormal.	5
14	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ and $Std_T < 740.0$ THEN class is abnormal.	5
15	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ , $740.0 \leq Std_T \leq 810.5$ and $1831.0 < Max_{R2} < 1840.0$ THEN class is abnormal.	6
16	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ , $740.0 \leq Std_T \leq 810.5$ , $Max_{R2} \leq 1831.0$ and $Avg_{R2} < 1693.0$ THEN class is abnormal.	7
17	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ , $740.0 \leq Std_T \leq 810.5$ , $Max_{R2} \leq 1831.0$ and $Avg_{R2} \geq 1693.0$ THEN class is normal.	7
18	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ , $740.0 \leq Std_T \leq 810.5$ , $Max_{R2} \geq 1840.0$ and $Avg_{R4} < 2626.0$ THEN class is abnormal.	7
19	IF $Avg_T \geq 2035.0$ , $Dur_T \leq 366.0$ , $2819.0 \leq Max_{R4} \leq 3006.0$ , $716.0 \leq Avg_{R6} \leq 804.0$ , $740.0 \leq Std_T \leq 810.5$ , $Max_{R2} \geq 1840.0$ and $Avg_{R2} \geq 2626.0$ THEN class is normal.	7
규칙 수	19	

( $Avg_T$ ,  $Dur_{R2}$ ,  $Avg_{R2}$ ,  $Avg_{R4}$ ,  $Dur_T$ ,  $Max_{R4}$ ,  $Max_{R2}$ ,  $Avg_{R6}$ ,  $Std_T$ ,  $Max_{R2}$ )이 선택되었다. 구축된 두 개의 의사결정 트리들에서 도출된 “if-then” 분류 규칙들의 집합은 각각 <표 3>, <표 4>에 정리하였다.

두 의사결정 트리의 복잡성과 분류 정확성을 비교한 결과는 <표 5>와 같다. 단, I종 오류는 정상을 이상으로 분류하는 경우이고 II종 오류는 그 반대의 경우를 의미한다.

<표 5>에서 보듯이 특징 선택에 의해 구축된 모델이 더 우수한 결과를 보였다. 특징 선택을 하지 않을 경우 분류 규칙의 수가 매우 많이 도출되었으며 if절에 포함된 평균 결정 노드 수 역시 특징 선택을 했을 경우가 더욱 적었다. 이는 if-then 규칙의 원인 절(Cause clause)이 복잡하여 이를 이용한 인과관계 분석이 힘들 수 있는 가능성을 내포한다. 그리고 정확성 측면에서 역시

정의된 모든 특징을 그대로 사용했을 경우 분류 에러가 3배나 높았다. 따라서 특징 선택 과정을 통해 공정에 대해 설명력 높

Table 5. 두 의사결정 트리의 성능 비교

		특징 선택 적용	특징 선택 미적용
복잡성	분류 규칙 수	10	19
	최대 if절에 포함된 결정 노드 수	7	7
	if절에 포함된 평균 결정 노드 수	3.9	4
분류 정확성	I종 오류	0%	0%
	II종 오류	0.83%	2.49%

은 특징들을 선택하고 이를 사용하여 FDC 모델을 구축하는 것은 모델의 복잡성을 낮출 수 있을 뿐만 아니라 불필요한 특징의 제거를 통해 모델의 정확성까지 향상시킬 수 있다.

#### 4. 결 론

본 논문에서는 반도체 제조 공정의 이상 진단과 분류를 위해 공정 설비에 부착된 센서로부터 습득되는 데이터 스트림을 이용하여 의사결정 트리 기반의 FDC 모델을 구축하였다. 모델 구축을 위해 데이터 스트림의 다양한 구조적 특징을 정의하고 이들 중 공정 결과에 대해 설명력 높은 요소들을 특징 선택 과정을 통해 추출하였으며 ID3 알고리즘을 이용하여 선택된 특징들에 의한 의사결정 트리 모델을 구축하였다. 기존 연구와 달리 본 연구에서는 데이터 스트림을 여러 개의 세부 영역들로 나누고 각각의 영역들과 전체 데이터 스트림으로부터 다양한 특징들을 정의하고 이를 이용한 FDC 모델을 구축했다. 구조적 패턴으로부터 얻어진 특징들은 공정의 상태 및 이상 상황을 더욱 직관적이고 자세하게 모니터링 할 수 있다. 또한 실험결과를 통해 알 수 있듯이, 특징 선택 과정을 통해 생성된 의사결정 트리는 그렇지 않은 모델에 비해 복잡성이나 정확성 모든 면에서 우수한 특성을 가졌으며 테스트 데이터 집합에 대해 뛰어난 이상 진단과 분류 성능을 보여주었다. 그리고 무엇보다 구축된 의사결정 트리를 통해, 단순히 공정의 결과를 제시하는데 그치지 않고 세부적인 원인의 분석이 가능하도록 하였다.

향후 본 논문의 제안 기법들이 실제 현업에 적용되기 위해서는 제안한 특징 선택 방법을 센서 선택으로 확장하여 단일 센서 차원의 연구를 공정내 주요 센서 그룹을 기준으로 도출된 특징들을 기반으로 하는 다변량 모델로의 확장해야 한다. 또한 설비의 노후화(Aging), 화학적인 오염, 윈도우 클라우딩(Window clouding), 그리고 환경적 변화 등의 요소로 인해 공정의 계측값이 서서히 증가하거나 감소하는 공정 드리프트(Process drift) 현상에 대해서도 능동적으로 반응하여 필요한 시점에 FDC 모델을 갱신하는 방안도 고려해야 한다. 이를 통해 공

정 드리프트 현상에 의해 발생하는 분류 에러를 최소화하고 중요 센서 그룹을 통한 더욱 신뢰성 있는 FDC 모델 구축이 가능하다.

#### 참고문헌

- Baek, J. G., Kim, C. O., and Kim, S. S. (2002), Online learning of the cause-and-effect knowledge of a manufacturing process, *International Journal of Production Research*, 40(14), 3275-3290.
- Cherry, G. and Qin, S. J. (2006), Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis, *IEEE Transactions on Semiconductor Manufacturing*, 19(2), 159-172.
- Choi, J. Y., Ko, J. M., Kim, C. O., Kang, Y. S., and Lee, S. J. (2007), Process start/end event detection and dynamic time warping algorithms for run-by-run process fault detection, In Proceedings of International Symposium on Semiconductor Manufacturing, 1-4.
- Dunia, R., Qin, S. J., Edgar, T. F., and McAvoy, T. J. (1996), Identification of faulty sensors using principal component analysis, *AIChE Journal*, 42(10), 2797-2812.
- Fayyad, U. M. and Irani, K. B. (1993), Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning, In Proceedings of the 13th International Joint conference on Artificial Intelligence, 1022-1027
- Goodlin, B. E., Boning, D. S., Sawin, H. H., and Wise, B. M. (2003), Simultaneous Fault Detection and Classification for Semiconductor Manufacturing Tools, *Journal of Electrochemical Society*, 150(12), 778-784.
- Guo, H.-F., Spanos, C. J., and Miller, A. J. (1991), Real time statistical process control for plasma etching, In Proceedings of IEEE/SEMI International Semiconductor Manufacturing Science Symposium, 113-118.
- Guyon, I. and Elisseeff, A. (2003), An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3, 1157-1182.
- Irani, K. B., Cheng, J., Fayyad, U. M., and Qian, Z. (1993), Applying machine learning to semiconductor manufacturing, *IEEE Intelligent Systems and Their Applications*, 8(1), 41-47.
- Kohavi, R. and John, G. (1997), Wrapper for feature subset selection, *Artificial Intelligence*, 19(1-2), 273-324.
- Lada, E. E., Lu, J.-C., and Wilson, J. R. (2002), A Wavelet-Based Procedure for Process Fault, *IEEE Transactions on Semiconductor Manufacturing*, 15(1), 79-90.
- Quinlan, J. R. (1986), Induction of decision trees, *Machine Learning*, 1(1), 81-106.
- Salvador, S., Chan, P., Brodie, J. (2004), Learning States and Rules for Time Series Anomaly Detection. In Proceedings of the 17th International Florida Artificial Intelligence Research Symposium.



**손지훈**

공군사관학교 전산통계학과 학사  
 연세대학교 정보산업공학과 석사  
 현재: 공군 제 6전대 235 비행대대 조종사  
 관심분야: 실시간 데이터 마이닝, RFID 기반  
 물류정보시스템

**고종명**

연세대학교 컴퓨터산업공학과 학사  
 연세대학교 컴퓨터산업공학과 석사  
 현재: 연세대학교 정보산업공학과 박사과정  
 관심분야: 실시간 데이터 마이닝, SCM,  
 Fault Detection and Classification

**김창욱**

고려대학교 산업공학과 학사  
 고려대학교 산업공학과 석사  
 Purdue University 산업공학과 박사  
 현재: 연세대학교 정보산업공학과 부교수  
 관심분야: 정보시스템 설계, RFID기반 물류  
 시스템, 실시간 데이터 마이닝