

고등지능정보공학II Take Home Exam

1번 회귀 문제 (50점)

1. 데이터 개요

반도체 제조에서 TSV 에칭(Through Silicon Via Etch)공정은 웨이퍼 기판에 깊은 마이크로 홀을 만드는 공정(그림 1)으로, 수십회의 에칭 작업을 반복하여 홀을 점진적으로 형성해 나간다. 1회 에칭 사이클은 홀을 뚫는 Etch step과 벽면을 안정화하는 Passivation step으로 구성되고 이런 사이클을 반복한다(그림 2).

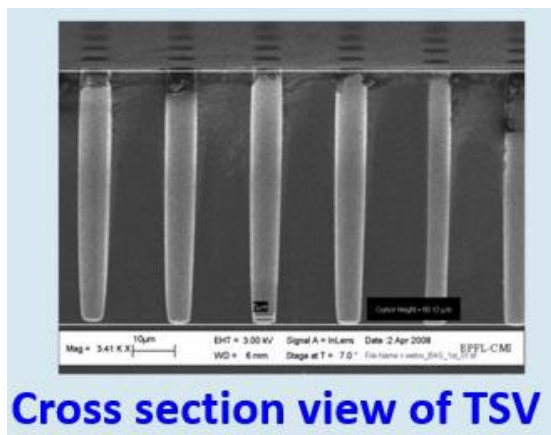


그림 1

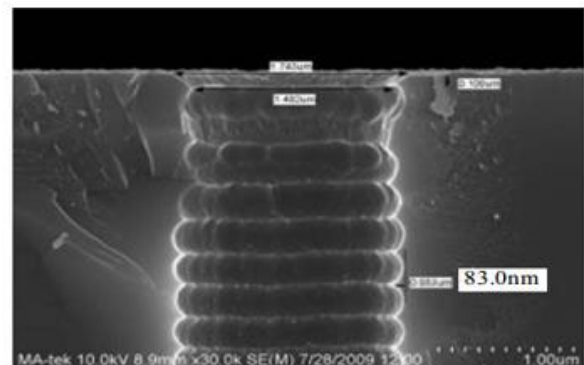


그림 2

아래 그림 3은 에칭 설비와 광학장비인 OES(Optical Emission Spectroscopy)를 보여준다. OES 장비는 에칭 공정이 진행되는 동안 산출물 원소의 에너지를 파장 형태로 기록하는 장비로 원소별로 특화된 파장의 에너지(intensity)를 가공시간동안 모니터링한다(그림 4).

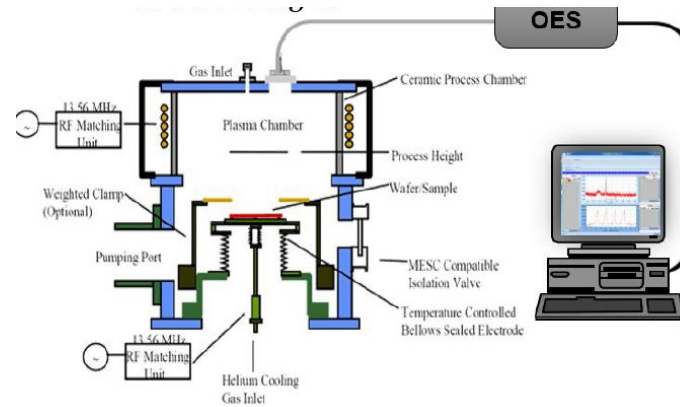


그림 3

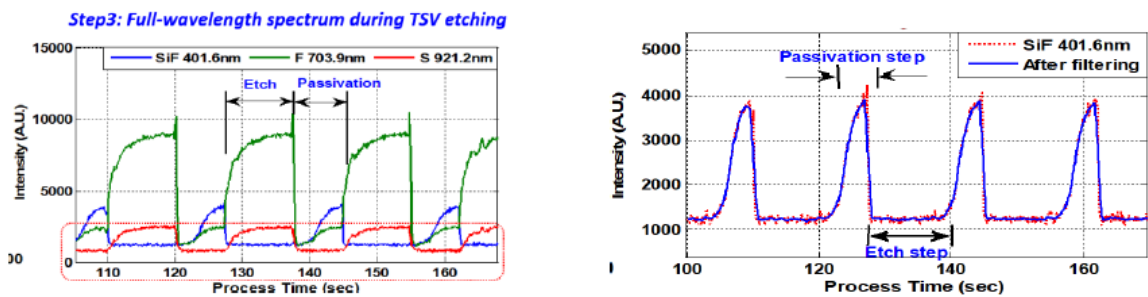


그림 4

웨이퍼 에칭동안 수집되는 파장의 스펙트럼은 162.969nm – 1017.36nm (3,648 indices)이며, 50회의 사이클을 반복한다. 데이터셋은 두개로 직경 80um와 25um 홀을 만든 결과이며 각 데이터셋마다 17개 웨이퍼의 OES 데이터(X에 해당)와 홀의 깊이(Depth로 y에 해당)가 존재한다. 또한 각 웨이퍼는 OES 데이터 뿐만 아니라 공정을 수행할 때 투입되는 레시피 파라미터를 모니터링한 데이터 (FDC Data로 X에 해당)도 있다. 한편 참고로 17개의 웨이퍼 중에서 12개는 정상적으로 진행된 데이터이며 5개는 비정상적으로 진행된 데이터이다(그림 5).

• Etch Recipe (normal Process)

Parameter	Etch	Passivation
Process Cycles	50	50
Time (sec)	11	8
SF ₆ (sccm)	130	-
O ₂ (sccm)	13	-
C ₄ F ₈ (sccm)	-	90
Source Power (W)	800	800
Platen Power (W)	20	0
Pressure (mTorr)	17	17

• Etch Recipe (Abnormal Process)

• 5 wafers each TSV diameter was etched (Fault Data)

Parameter	#Sample	Adjustment
Process Cycles	All	50
C ₄ F ₈ (sccm)	1	90 → 85
O ₂ (sccm)	2	13 → 11
SF ₆ (sccm)	3	130 → 122
Source Power (W)	4	800 → 760
Platen Power (W)	5	17 → 15

그림 5

2. 데이터 정보

데이터를 정리하면 다음과 같다.

데이터셋 1 (25 μ m)	12개의 정상 데이터, 5개의 비정상 데이터 X: OES와 FDC Data (모두 시계열 형식), y: 공정결과인 홀의 깊이
데이터셋 2 (85 μ m)	12개의 정상 데이터, 5개의 비정상 데이터 X: OES와 FDC Data (모두 시계열 형식), y: 공정결과인 홀의 깊이

3. 문제

이 문제는 대표적인 “large p, small n문제”로, 각 데이터셋에 대해서 다음 과제를 수행하시오.

(1) 3차원 데이터셋(샘플수x변수x가공시간)을 2차원으로 unfolding한 후 일반적인 PCA와 kernel PCA를 수행하고 PCR 모델의 성능을 leave-one-out cross validation(총 17번)을 통해 측정하시오 (R2, MAE, MAPE) (25점)

(2) 여러분이 알고 있는 다른 예측 모델을 이용해 X에 대해 y를 예측하는 회귀 모델을 제시하고, 모델의 성능을 leave-one-out cross validation(총 17번)을 통해 측정하시오 (R2, MAE, MAPE) (25점)

*MAE : Mean Squared Error

MAPE : Mean Absolute Percentage Error

*성능 지표의 경우 cross validation를 17번 수행하여 얻은 모든 샘플에 대한 예측값을 모아 한 번에 계산

2번 분류 문제 (35점)

1. 데이터 개요

제조 공정에서 가공품(웨이퍼)의 양/불을 판정하는 이진 분류 문제이다. predictors는 435개이고 response는 정상(0) 또는 불량(1)이다. Predictors는 공정에서 모니터링된 센서 값이며 이 센서 값을 이용해서 공정이 완료된 후 가공품이 정상인지 불량인지를 판정한다. 총 학습 데이터 수는 426개이다. 이 문제의 특징은 데이터 수보다 독립변수 수가 더 많은 고차원 데이터(high-dimensional data)를 분석해야 하는 것이며, 정상과 불량 데이터 수가 크게 차이나지 않는 클래스 균형(class balance) 문제이다.

2. 문제

학습 데이터(X_train.csv)를 활용하여 모델을 학습하고 학습된 모델로 테스트 데이터(X_test.csv)에 대해 제품의 불량 여부를 분류한 결과를 <sample_submission.csv 파일 생성 예시>를 참고하여 제출하시오. 성능을 높이기 위해 사용한 기법이나, 사용한 이유 등에 대해 자세히 설명하시오. (제출한 sample_submission.csv 파일은 F1 score 기준으로 평가) (35점)

<sample_submission.csv 파일 생성 예시>

predict
1
1
...
...
-1

3번 이론 문제 (15점)

1. 문제

Multidimensional Scaling 알고리즘을 공부하여 자세히 논리적인 구조로 서술하시오(Metric 모델만 설명)

제출기한

11월 3일 23시 59분 (LearnUS에 보고서와 코드를 업로드하시오.)

유의사항

1번 문제의 경우 보고서 내에 예측 성능을 표로 정리하고, 분석에 대한 논리적 이유를 최대한 상세히 작성하시오.

2번 문제의 경우 예측 파일을 정해진 양식대로 제출하고, 보고서 내에 분석에 대한 논리적 이유를 최대한 상세히 작성하시오.

3번 문제의 경우 보고서 내에 작성하시오.

즉, 제출 파일은 보고서 파일과 예측 파일임

보고서에는 문제에 대한 답변 뿐만 아니라 수행한 분석에 대한 논리적인 이유가 있어야 함.

다른 학생의 코드를 일부, 전부 표절한 경우 두 학생 모두 0점 처리함.

늦게 제출할 경우 10분당 5점씩 감점함.