**[제목]** 시계열 데이터 패턴 비교 알고리즘 (DTW) 최적화

**[내용]**

  - 패턴 비교 알고리즘 (DTW) 최적화

    .반도체 시계열 데이터의 패턴을 분석하기 위해선 다양한 패턴 비교 알고리즘들을 활용하고 있음

    .DTW (Dynamic Time Warping)은 전통적인 패턴 비교 알고리즘 중 하나로, DTW 내에서도 연산속도 개선을 위한 다양한 DTW 알고리즘이 존재

  - 본 과제에서는 다양한 DTW 알고리즘 중 time complexity를 최소로 하는 알고리즘 찾고 구현하는 것을 목표로 함

  1) 알고리즘에 활용 가능한 반도체 센서 시계열 데이터 DB Schema 설계

  2) 다양한 DTW 알고리즘 조사 및 구현

  3) Time Complexity 기반 성능 평가

**[필요 지식]**대용량 데이터 처리, 프로그래밍 스킬(C#, C++), 패턴 비교 알고리즘 지식

**[교육/훈련 효과]** DB, 데이터 분석 알고리즘 Study, 프로그래밍 기법 습득

**[개발 기간]** 3개월

**[개발 인원]**3인

□ 과제 개요 #2

**[제목]** 센서 특성 자동 분류 알고리즘 개발

**[내용]**

  - 반도체 센서 데이터의 복잡성

    .설비 모델 別 센서들은 다양한 물리적 특성을 보유하고 있음 (ex. 1) Step 구분 필요 / Step 구분 불필요, 2)

    .기존 방식에선 설비 명, 센서 명 및 시계열 데이터를 기반으로 수동으로 이를 구분하고 업데이트 하는 작업들을 수행

  - 본 과제에서는 센서의 시계열 데이터를 활용 자동적으로 센서들의 특징을 분류해 줄 수 있는 알고리즘 개발을 목적으로 함

    (ex. Decision Tree, Support Vector Machine, Deep learning: CNN/RNN 등)

  1) 알고리즘에 활용 가능한 반도체 센서 데이터 DB Schema 설계

  2) 시계열 데이터 기반 특성 추출 알고리즘 개발

  3) 센서 특성 자동 분류 알고리즘 개발

**[필요 지식]**대용량 데이터 처리, 프로그래밍 스킬(언어 무관), 인공지능 알고리즘

**[교육/훈련 효과]** DB, 딥러닝 관련 라이브러리 활용 경험, 프로그래밍 기법 습득

**[개발 기간]** 3개월

**[개발 인원]**3인

□ 과제 개요 #3

**[제목]** 반도체 생산 설비간 Input parameter 유사도를 판단하기 위한 criteria 정의 및 계산 알고리즘 개발

**[내용]**

  - 반도체 생산 설비 Input parameter의 복잡성

    . 특정 공정의 경우, 설비당 약 3만개 ~ 40만개의 Input parameter 존재.

    . 설비 모델별로 parameter 값이 같아야 하는 것과, 일정 분포 안에 있어야 하는 것이 있으나,

      parameter 별로 명확히 정의되어있지 않음. (설비별 설정되어 있는 값만 존재)

  - 본 과제에서는, 실제 반도체 생산 설비에 설정되어 있는 Input parameter 데이터를 바탕으로 (보안상 데이터를 변형하여 일부 제공할 예정)

**Classification**- 각 parameter를 설비별로 같아야 하는 것과 k개의 cluster를 가지는 것, 일정 분포 안에 있어야 하는 것으로 분리

**Abnormaly detection**- 오설정으로 판단되는 parameter 검출

    을 위한 criteria 정의 및 계산 알고리즘을 도출하는 것을 목적으로 함.

  1) 알고리즘에 활용 가능한 반도체 생산 설비 input parameter DB 설계

  2) 설비간 일치도를 판단하기 위한 criteria 정의

  3) 데이터 분석 algorithm 개발

**[필요 지식]**대용량 데이터 처리, 프로그래밍 스킬(언어 무관), Data mining

**[교육/훈련 효과]** DB, 프로그래밍 기법 습득, Data mining 기법 활용

**[개발 기간]** 3개월

**[개발 인원]**3인

□ 과제 개요 #4

**[제목]** Semi-supervised Learning 기반 공정 결과 예측 알고리즘 개발

**[내용]**

  - 반도체 공정 데이터의 풍요속의 빈곤

    . 데이터는 풍부하나 Label된 data는 부족함

    . 전수 계측이 불가능하기 때문에 Sampling 계측 수행 (2 wafers / 2 lot (50 wafers))

    . 양산 수율 안정화로 인한 불량 데이터 발생 빈도 저조 → 양품 대비 불량 비율이 극히 낮음

  - 본 과제에서는 Sampling 되어 계측된 데이터 (Label 有)와 계측되지 않은 wafer의 설비 데이터 (Label 無)를 활용한

    Semi supervised Learning 기법을 활용하여 공정 결과 예측하는 알고리즘 개발을 목표로 함

  1) 시계열 데이터 기반 학습 알고리즘 개발

  2) Semi supervised learning 기반 센서 데이터와 품질간의 연관성 도출 알고리즘 개발

  -

**[필요 지식]**대용량 데이터 처리, 프로그래밍 스킬(언어 무관), 인공지능 알고리즘

**[교육/훈련 효과]** DB, 딥러닝 관련 라이브러리 활용 경험, 프로그래밍 기법 습득

**[개발 기간]** 3개월

**[개발 인원]**3인