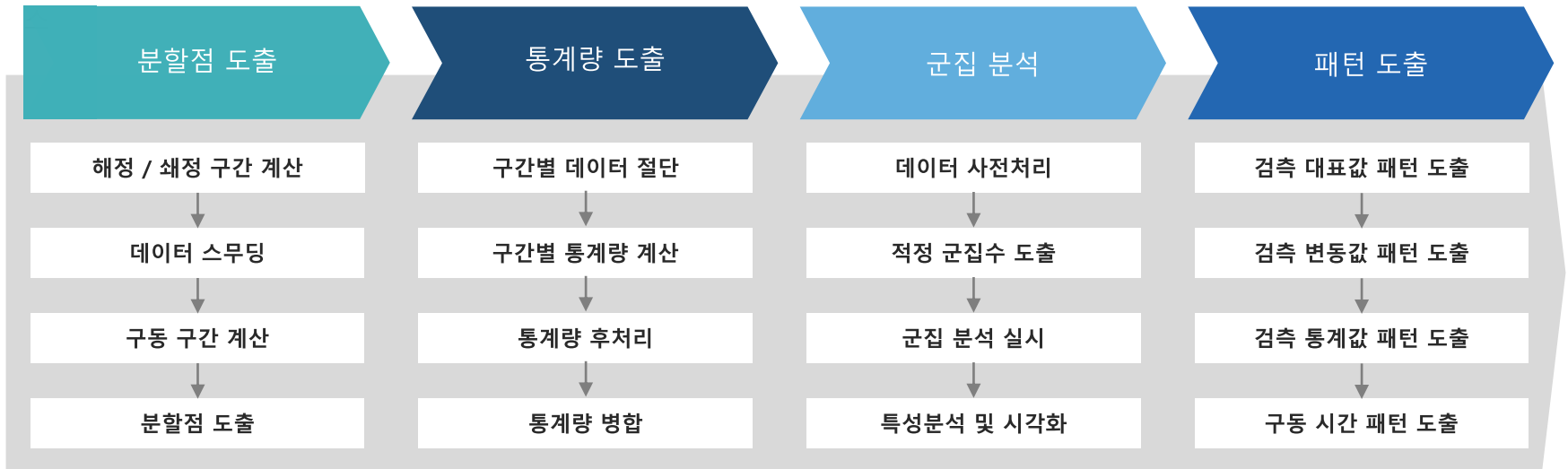


- 선로전환기 패턴 도출 정의를 통해 구간별 패턴 분석 결과 도출
- 분할점 도출, 통계량 도출, 군집 분석, 패턴 분석 등 총 4단계 수행을 통하여 선로전환기의 패턴 도출

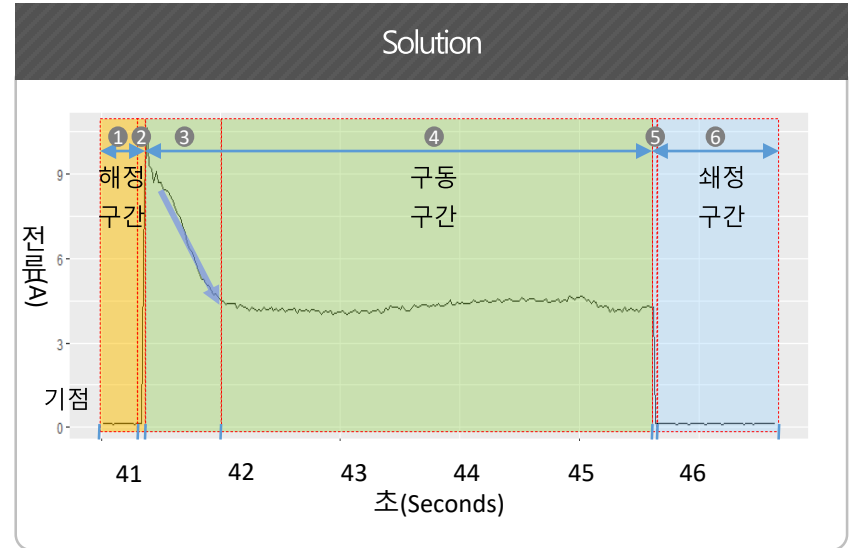
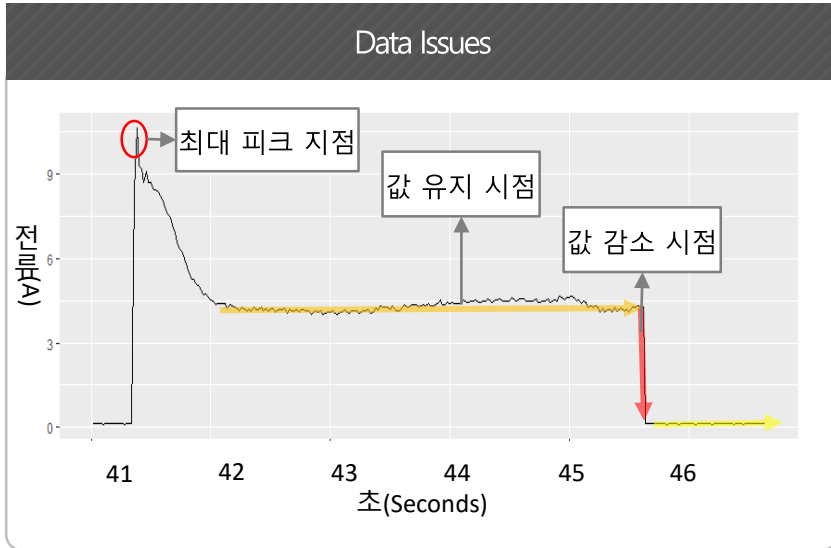
✓ 선로전환기 패턴 분석 프로세



✓ 선로전환기 패턴 정의

번호	군집 유형	군집 내용	순번	군집 유형	군집 내용
1	군집 번호 1	Classified	5	군집 번호 5	Classified
2	군집 번호 2		6	군집 번호 6	
3	군집 번호 3		7	군집 번호 7	
4	군집 번호 4		8	군집 번호 8	

- 탐색적 데이터 분석(EDA) 과정에서 선로전환기의 구간별 특성을 확인
- 구간별 분할을 통해 각 구간별 특성을 독립적으로 분석하여 구간별 특성 차이를 패턴 분석에 활용



구간별 유형 분류 필요

- ✓ 선로전환기 데이터의 유형 분류를 위하여 유형별 데이터의 특성에 대한 분류 필요 확인
- ✓ 선로전환기 데이터 탐색을 통하여 구간의 유형별 특성에 대한 구분이 가능함을 확인

* 구간별 특성 : 최대 피크지점이 존재하거나 최대 피크 이후 값이 감소하는 지점, 값이 유지되는 지점, 값이 급격히 떨어지는 지점

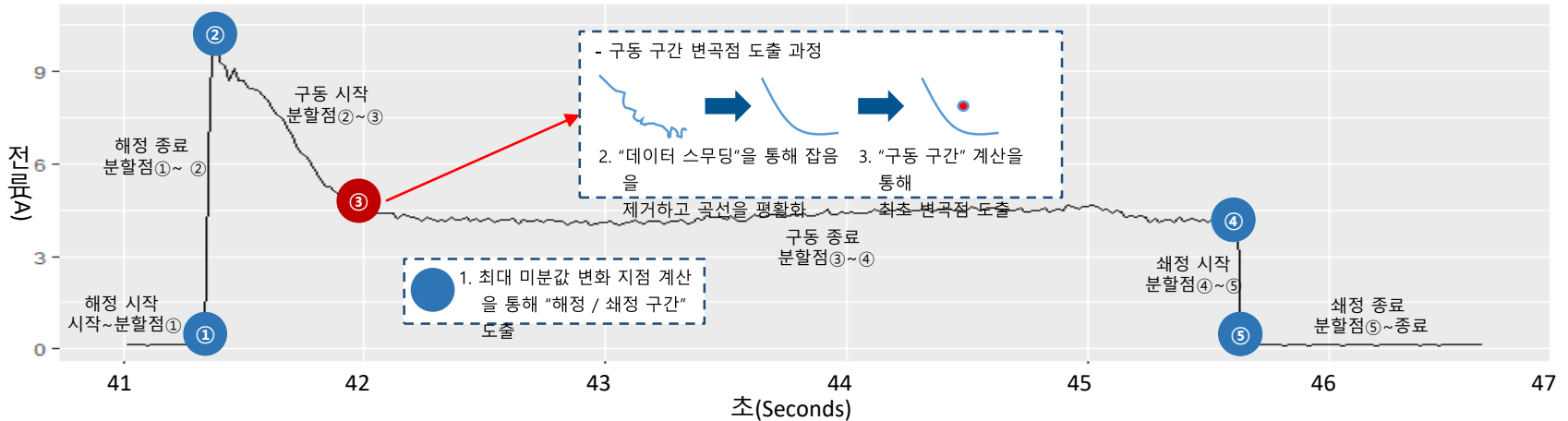
구간별 분할 실시

- ✓ 현업 및 전문가와 소통 결과 각 구간별 특징이 뚜렷하게 발현되는 지점을 해정 / 구동 / 쇄정 구간으로 분할 실시
- ✓ 데이터를 다음과 같은 의미를 갖는 6개 구간 분할

구간 구분	구간 번호	구간 내용
해정 시작 / 종료	1) 기점-①시작 2) ①시작-②종료	선로전환기의 최초 관측 시작부터 피크 지점까지 구간
구동 시작 / 종료	3) ②시작-③종료 4) ③시작-④종료	선로전환기 검측값의 피크 이후 값의 하락부터 유지 지점까지 구간
쇄정 시작 / 종료	5) ④종료-⑤시작 6) ⑤시작-⑥종료	선로전환기의 검측 값의 감소 이후 관측 종료 지점까지 구간

- 수학적, 통계적 근거에 기반하여 구간을 분할하는 구간 분할 알고리즘에 대한 설명
- 5개의 분할점을 모든 이벤트에 대하여 자동으로 도출하여 구간별 통계량 생산에 활용

구간 분할 알고리즘 작성



1. 해정 / 쇄정 구간 계산

- ✓ 해정 시작/ 해정 종료/ 쇄정 시작/ 쇄정 종료 구간의 시작과 종료 지점 네 곳은 수학적으로 2차 미분한 값이 모든 데이터 중 가장 큰 지점들
- ✓ 이 지점들의 위치를 도출하여 분할 지점으로 선정

2. 데이터 스무딩

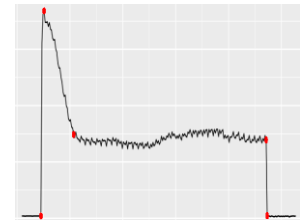
- ✓ 데이터 스무딩 : 잡음 등의 이유로 매끄럽지 않은 곡선을 매끄럽게 만들어 주는 방법론
- ✓ 구동 종료 구간 계산을 위해 데이터를 임시적으로 스무딩함

3. 구동 구간 계산

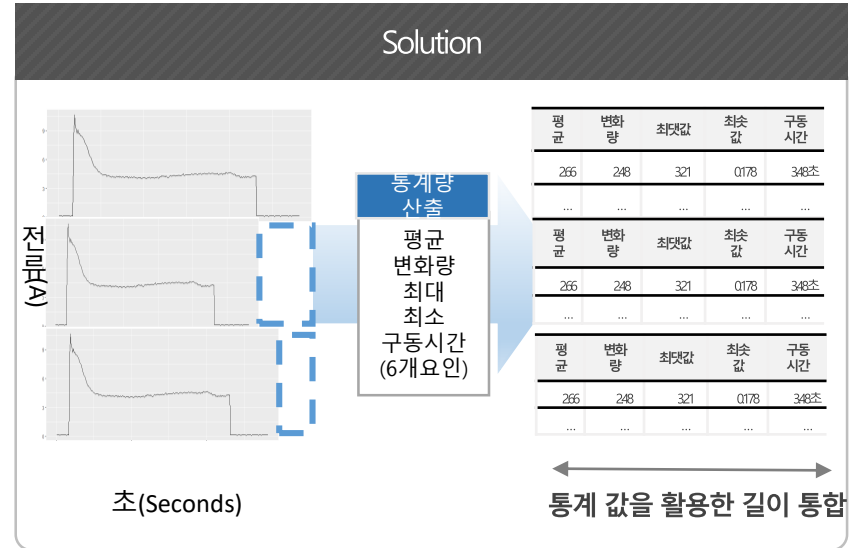
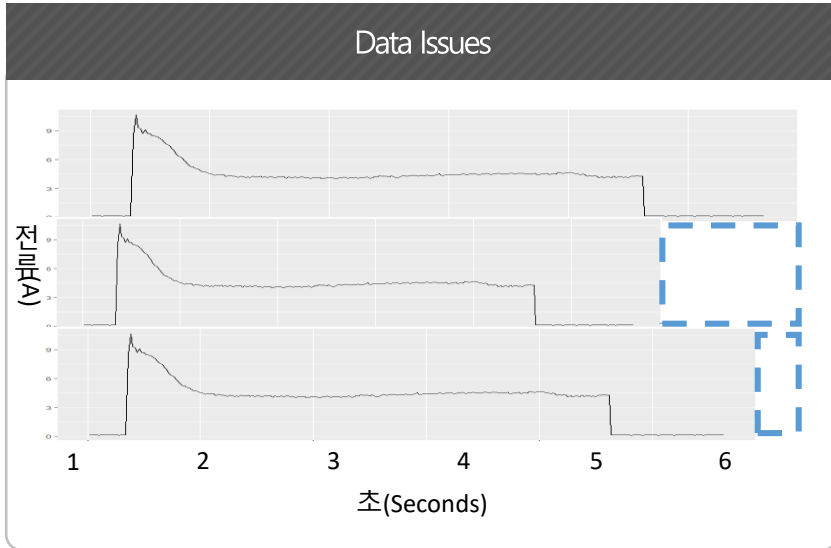
- ✓ 구동 시작 구간은 기울기의 변화가 완만하여 2차 미분값만으로 도출이 어려움
- ✓ 스무딩 이후 2 ~ 4번째 구간을 절단 후, 최초 변곡점을 구동 시작 구간으로 지정

4. 분할점 도출

- ✓ 5개 분할점 최종 도출
- ✓ 구간별 통계량 생산에 활용



- 탐색적 데이터 분석(EDA) 과정에서 이벤트 발생 마다 시간 길이가 차이남을 확인
- 각 구간별 통계량 변환을 실시하여 데이터의 열 길이를 일치



. 데이터 시간 길이 불일치

- ✓ 선로전환기 구동 이벤트에 대한 시간 길이 이슈 확인
- ✓ 시간 길이의 차이는 분석에 대한 동일 조건이 충족되지 않기 때문에 다양한 분석 기법에 대한 적용의 어려움 확인
- ✓ 다양한 분석 방법론 적용을 통하여 데이터 시간 길이의 일치를 수행하여 각 구동 이벤트별 분석 요건을 충족시킬 필요 확인

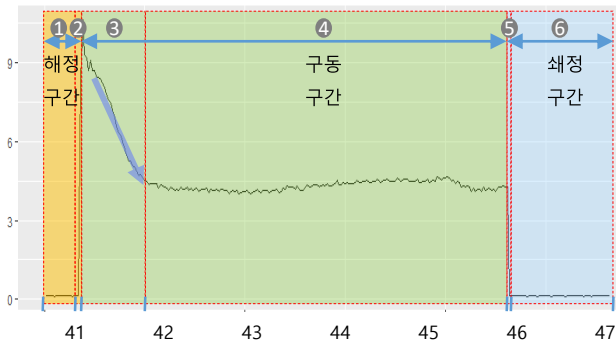
구간별 통계량 변환

- ✓ 선로전환기 구간 분할 알고리즘으로 도출한 6개 구간에 대하여 각각 통계량 변환을 실시
- ✓ 구간별 통계량 지표를 활용하여 각 구간에 대한 특징 산출 수행
- ✓ 평균, 변화량, 최대, 최소, 시간 길이 등 각 구간별 특징에 대한 통계량을 선정 및 변환 작업에 적용

- 도출된 구간 분할점별로 구간을 절단하여 평균, 변화량, 최댓값, 최솟값, 시간 길이 통계량을 생산
- 데이터의 시간 불일치 이슈를 해소하고 데이터의 특성을 뚜렷하게 드러내는 효과 기대

구간별 통계량 생성

1 구간별 데이터 절단



2 분할점 이상 예외 처리

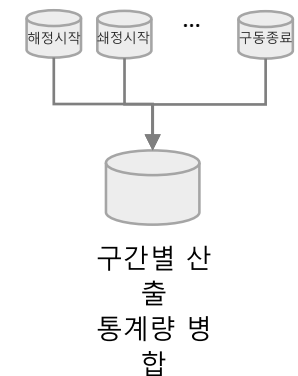
- ✓ 구간별 분할점 예외 사항 확인을 통한 처리

Classified

3 구간별 통계량 계산

구간	평균	변화량	최댓값	최솟값	구동 시간
해정 시작	01	01	005	009	011
해정 종료	569	997	6	01	1083
구동 시작	723	749	21	403	1083
구동 종료	378	376	08	342	422
쇄정 시작	187	18	18	011	37
쇄정 종료	01	01	005	009	012

4 구간별 통계량 병합



1. 분할점 적용을 통한 구간화

- ✓ 분할점 알고리즘으로 도출한 5개의 분할점을 데이터에 적용
- ✓ 데이터를 해정 시작 / 해정 종료 / 구동 시작 / 구동 종료 / 쇄정 시작 / 쇄정 종로의 6 구간으로 절단

2. 분할점 예외 처리

- ✓ 선로전환기의 구간별 분할점 알고리즘을 벗어난 이벤트 도출
- ✓ 해당 이벤트들에 대한 예외 사항으로 정의 수행

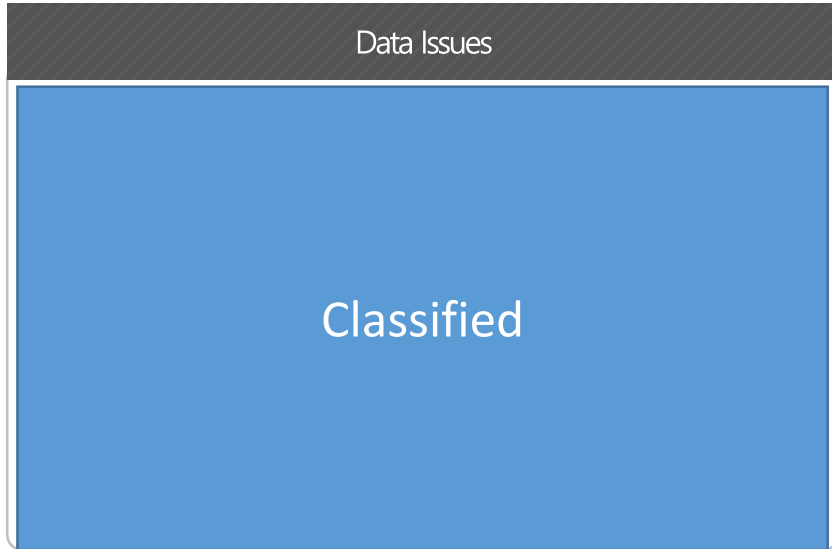
3. 구간별 통계량 계산

- ✓ 도출한 6개 구간에 대하여 각각 5개의 통계량을 계산
- ✓ 구간별 도출 통계 지표
 - . 평균 : 구간의 평균
 - . 변화량 : 구간의 변화폭
 - . 최댓값 : 구간의 최댓값
 - . 최솟값 : 구간의 최솟값
 - . 시간 길이 : 구간 시간 길이

4. 통계량 병합

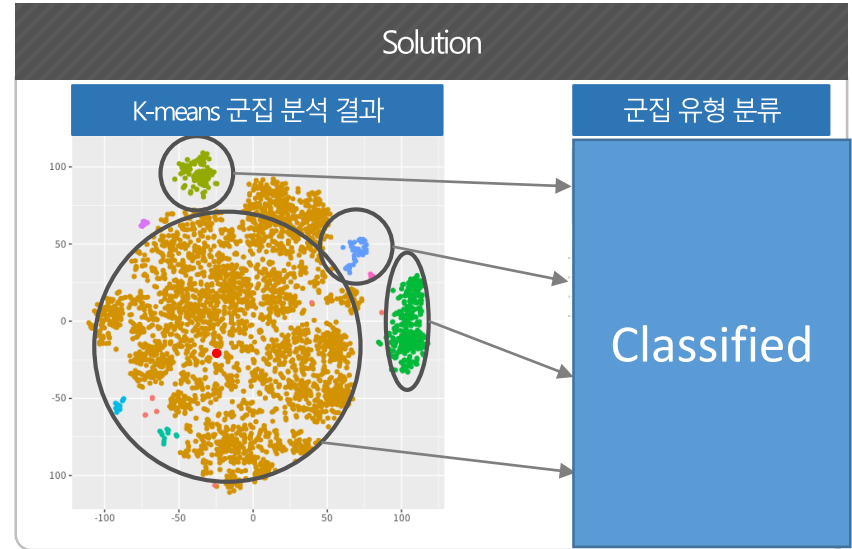
- ✓ 6개 구간에 대하여 계산한 5개의 통계량으로 하나의 이벤트당 30개의 통계량이 생산
- ✓ 전체 49,741개의 이벤트에 대하여 30개의 통계량을 계산하여 하나로 병합

- 탐색적 데이터 분석(EDA) 과정에서 여러 패턴에 대한 유형 확인
- K-means 군집 분석을 통하여 선로전환기 전체 이벤트에 대하여 패턴 유형 분류 결과 도출



패턴 유형 분류 불가

- ✓ 선로 전환기의 모든 구동 이벤트 49,741개 중엔 다양한 패턴을 그리는 여러 유형의 데이터가 섞여 있음
- ✓ 패턴 분석을 수행하기 위해선 섞여 있는 데이터를 여러 유형의 독립된 집단으로 분할하여 각각의 특징을 분석해야 함
- ✓ 유형을 분할하는 적절한 기준값이 존재하지 않는 제약 상황에서 유형 분리를 수행하는 분석적 방법론이 요구됨



K-means 군집 분석

- ✓ 사전 제약 없이 유형 분리를 수행하는 분석적 방법론으로 K-means 군집분석을 적용 가능함
- ✓ K-means 군집분석은 데이터들의 값 간 거리를 계산하여 가까이 있는 데이터들은 한 군집으로 묶어주고, 멀리 있는 데이터들은 다른 군집으로 묶는 알고리즘적 방법론을 의미함
- ✓ K-means 군집분석을 통해 데이터 패턴들을 유사한 특성을 공유하는 여러 군집들로 묶는 것이 가능

- 각 구간별 통계량을 기반으로 산출된 패턴들을 각 군집으로 분리하여 특성 및 시각화 수행
- 이벤트에 대한 패턴이 정의를 위하여 K-means 군집 분석 방법론을 적용하여 군집 유형 분류 수행

군집분석 프로세스

