

WEEK 10 과제

<기계학습 구현>

2021254019 김지현

목차

1. 데이터의 종류
2. 데이터의 특성
3. 분석 목적
4. 선택한 기계학습 알고리즘

1. 데이터 종류

- ▶ 실제 현장에서 얻을 수 있는 데이터 중에서 선택하였음.
- ▶ 선택 데이터
 - ▶ 전기 배터리 **Cell**의 외관불량 검사를 하는 설비를 개발 중에 있음.
 - ▶ 설비에서 매일 날짜마다 **csv**파일로 나오는 결과 **Data** 수집
 - ▶ 데이터 항목 : 광학계 비전 항목, 불량 검사항목(약 **70개** 이상), 각 불량 항목별 **Reject Count** 개수

TOTAL COUNT		16319
PASS COUNT		12184 74.661
REJECT COUNT		2702 16.557
INVALID COUNT		1433 8.781
V5	AA1_An_Stab	34
V10	AA1_An_Stab_DL	81
V10	AA1_Ca_Stab	70
V5	AA1_Ca_Stab	79
V5	AA1_Ca_TopStab	2
V1	AB1_Stab	382
V7	AB1_Stab	2204
V11	AB2_InnerStab	51
V9	AC1_BOTTOM_Pouch	89
V9	AC1_BOTTOM_Side	429
V4	AC1_BOTTOM_Stab_Pouch	2959
V4	AC1_BOTTOM_Stab_Side	3252
V8	AC1_Left_Stab_Side	27
V8	AC1_Right_stab_side	27
V7	AC1_Stab	418
V9	AC1_TOP_Dent_Pouch	307
V9	AC1_TOP_Dent_Side	45
V4	AC1_TOP_Stab_Pouch	1008
V4	AC1_TOP_Stab_Side	1019
V1	AD1_StabUpper	10
V10	BA1_An_Scratch	1
V5	BA1_An_Scratch	7
V10	BA1_Ca_Scratch	297
V5	BA1_Ca_Scratch	502
V1	BB1_Scratch	1044
V7	BB1_Scratch	627
V1	BD1_ScratchUpper	2
V6	CB3_Particle	1
V1	CB3_Side_Particle	160
V7	CB3_Side_Particle	335
V11	DB5_Elec_Line_Wrinkle	68

2021.04.16

TOTAL COUNT		7178
PASS COUNT		5903 82.237
REJECT COUNT		1076 14.99
INVALID COUNT		199 2.772
V5	AA1_An_Stab	31
V10	AA1_An_Stab_DL	76
V10	AA1_Ca_Stab	62
V5	AA1_Ca_Stab	54
V5	AA1_Ca_TopStab	2
V1	AB1_Stab	300
V7	AB1_Stab	1876
V11	AB2_InnerStab	47
V9	AC1_BOTTOM_Pouch	74
V9	AC1_BOTTOM_Side	330
V4	AC1_BOTTOM_Stab_Pouch	1793
V4	AC1_BOTTOM_Stab_Side	2053
V8	AC1_Left_Stab_Side	20
V8	AC1_Right_stab_side	23
V7	AC1_Stab	333
V9	AC1_TOP_Dent_Pouch	267
V9	AC1_TOP_Dent_Side	41
V4	AC1_TOP_Stab_Pouch	606
V4	AC1_TOP_Stab_Side	614
V1	AD1_StabUpper	8
V5	BA1_An_Scratch	7
V10	BA1_Ca_Scratch	266
V5	BA1_Ca_Scratch	437
V1	BB1_Scratch	866
V7	BB1_Scratch	507
V1	BD1_ScratchUpper	2
V6	CB3_Particle	1
V1	CB3_Side_Particle	144
V7	CB3_Side_Particle	279
V11	DB5_Elec_Line_Wrinkle	61
V5	EA1_An_Folding	4

2021.04.17

TOTAL COUNT		16334
PASS COUNT		13541 82.901
REJECT COUNT		2519 15.422
INVALID COUNT		274 1.677
V5	AA1_An_Stab	29
V10	AA1_An_Stab_DL	76
V10	AA1_Ca_Stab	61
V5	AA1_Ca_Stab	51
V5	AA1_Ca_TopStab	2
V1	AB1_Stab	276
V7	AB1_Stab	1781
V11	AB2_InnerStab	40
V9	AC1_BOTTOM_Pouch	70
V9	AC1_BOTTOM_Side	308
V4	AC1_BOTTOM_Stab_Pouch	1655
V4	AC1_BOTTOM_Stab_Side	1909
V8	AC1_Left_Stab_Side	20
V8	AC1_Right_stab_side	22
V7	AC1_Stab	316
V9	AC1_TOP_Dent_Pouch	250
V9	AC1_TOP_Dent_Side	38
V4	AC1_TOP_Stab_Pouch	558
V4	AC1_TOP_Stab_Side	564
V1	AD1_StabUpper	8
V5	BA1_An_Scratch	7
V10	BA1_Ca_Scratch	251
V5	BA1_Ca_Scratch	417
V1	BB1_Scratch	828
V7	BB1_Scratch	494
V1	BD1_ScratchUpper	2
V6	CB3_Particle	1
V1	CB3_Side_Particle	141
V7	CB3_Side_Particle	265
V11	DB5_Elec_Line_Wrinkle	60
V5	EA1_An_Folding	4

2021.04.18

2. 데이터 특성

- ▶ 학습에 사용하기 위해서는 **RAW**데이터에서 전처리 작업이 필요함
- ▶ 데이터 특성
 - ▶ 날짜순으로 쌓이는 데이터이다.
 - ▶ 각 검사항목별 **Reject Count**를 수집할 수 있다.
 - ▶ 1일전의 데이터와 **Count** 개수 차이를 계산하는 데이터 전처리 작업이 필요하다.
 - ▶ 정상적인 데이터를 **Good**으로 학습시키고, 그 범주에서 벗어나는 데이터를 **Bad**로 출력한다.

3. 분석 목적

- ▶ 정상적인 데이터와 소량의 비정상적인 데이터를 모아서 학습시키면, 특정 데이터를 집어넣었을 때, “몇일날에 어떤 항목이 비정상적이다”를 탐색해 내는 것이 목표이다.

- ▶ 예

<정상 데이터>

Date	Item	Count
2021년 04월02일	AB1_Stab	382
	BB1_Scratch	162
	FB1_Dent	88

Date	Item	Count
2021년 04월03일	AB1_Stab	298
	BB1_Scratch	210
	FB1_Dent	92

<비정상 데이터>

Date	Item	Count
2021년 04월05일	AB1_Stab	1985
	BB1_Scratch	162
	FB1_Dent	88

4. 선택한 기계학습 알고리즘

▶ 비지도형 기계학습

정답 레이블이 주어지지않고 데이터의 구조나 분포를 이해하여 데이터의 공통점 유무에 따라 반응하는 학습 방법이기 때문에 이 형식을 채택하였다.

비지도형 기계학습 의 여러 분류 중에서도 이상탐지 학습 방법을 생각해보았다.

이상탐지 기법은 단순히동평균이나 저대역 필터와 같은 통계 알고리즘을 활용하여 수행된다.

따라서 주어진 데이터셋에서 비정상적인 데이터를 찾을 것이라 기대되어 선택하게 되었다.

이를 수행하기 위해서는 데이터셋을 정상적인 예제와 비정상적인 예제를 준비하여 학습에 사용하여야 한다.