

1

사전테스트 문제

□ 다음 제시되는 데이터 정의서와 데이터를 활용하여 환경에 따른 딸기의 성장도를 예측하는 모델을 개발

○ 데이터셋

1) 충남 논산 지역 농가 4곳의 생육데이터 및 환경데이터

○ 평가 방법

1) INPUT : 환경데이터 / OUTPUT : 생육데이터

2) 생육데이터 데이터에 대한 정확도 판정

(정확도에 대한 판정은 MASE 지표를 활용)

○ 데이터셋 구성

1) 생육데이터

번호	항목	단위	주기	측정방법
1	초장	mm	1회/주	관부에서 가장 긴 잎의 선단까지의 길이
2	엽장	mm	1회/주	완전히 전개된 최근 3번째 잎을 대상으로 하며, 엽육이 시작되는 지점부터 잎 끝부분까지 길이
3	엽폭	mm	1회/주	적히 전개된 최근 3번째 잎을 대상으로 하며, 가장 폭이 넓은 부분 측정(엽장과 수직 방향이 되도록)
4	엽병장	mm	1회/주	관부에서 초장을 측정하는 잎의 하단까지의 길이(초장-엽장)
5	엽수	개	1회/주	한 개체의 잎의 수(완전히 전개된 잎만 측정) ※ 잎색이 연하거나 특별한 이상 증상 없이 잎표면이 주글거리면 미전개 잎으로 판단
6	관부직경	mm	1회/주	원줄기 관부에서 가장 굵은 부분의 길이
7	화방꽃수	개	1회/주	화방에 달린 꽃의 수
8	착과수	개	1회/주	과실의 수

2) 환경 데이터

번호	항목	변수명	주기	비고
1	공급EC	supplyEC	1시간	
2	공급PH	supplyPH	1시간	
3	내부CO2	innerCO2	1시간	
4	내부습도	innerHum	1시간	
5	내부온도	innerTemp	1시간	
6	내부일사량	innerSolar	1시간	

○ 작기정보

번호	농가	작기시작일	작기종료일	비고
1	B농가	20230909	20240424	
2	C농가	20230920	20240426	
3	D농가	20230920	20240424	
4	E농가	20230920	20240426	

성능평가 순서



□ 성능평가1 (정확도)

- 예측값에 대한 정확도를 MASE(Mean Absolute Scaled Error) 지표를 이용해 계산
- 성장도에 대한 정확도를 계산
- 수식은 아래와 같으며 Python numpy 라이브러리를 이용해 코드 구현 예정

$$MASE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|} \right)$$

□ 성능평가2 (속도)

- 예측에 소요된 시간을 전체 예측 대상 데이터셋의 수로 나누어, 하나의 데이터셋 예측에 소요되는 시간을 계산
- 사전 테스트의 경우에는 성장도 예측 속도를 계산
- 수식은 아래와 같으며, Python time 라이브러리를 소요시간 계산 예정

$$V = \frac{\times \text{tamp}_e - \times \text{tamp}_s}{n}$$

3

최종 성능평가 점수 계산

□ 최종 점수 계산식

○ 최종점수 = 정확도(80점) + 예측속도(20점)

순위	예측정확도		예측속도	
	점수	가중치 적용	점수	가중치 적용
1	100	80	100	20
2	98	78.4	98	19.6
3	96	76.8	96	19.2
4	94	75.2	94	18.8
5	92	73.6	92	18.4
6	90	72	90	18
7	88	70.4	88	17.6
8	86	68.8	86	17.2
9	84	67.2	84	16.8
10	82	65.6	82	16.4
11	80	64	80	16
12	78	62.4	78	15.6
13	76	60.8	76	15.2
14	74	59.2	74	14.8
15	72	57.6	72	14.4
16	70	56	70	14
17	68	54.4	68	13.6
18	66	52.8	66	13.2
19	64	51.2	64	12.8
20	62	49.6	62	12.4
21	60	48	60	12
22	58	46.4	58	11.6
23	56	44.8	56	11.2
24	54	43.2	54	10.8
25	52	41.6	52	10.4
26	50	40	50	10
27	48	38.4	48	9.6
28	46	36.8	46	9.2
29	44	35.2	44	8.8
30	42	33.6	42	8.4
31	40	32	40	8
32	38	30.4	38	7.6
33	36	28.8	36	7.2
34	34	27.2	34	6.8
35	32	25.6	32	6.4
36	30	24	30	6
37	28	22.4	28	5.6

○ 제출 파일

- 1) 학습이 완료된 모델 파일
- 2) 모델을 정의하고, 제시된 Input Data를 받아 Output Data를 추론하는 코드가 정의된 .py 파일 (inference.py)

2-1) Input

- csv 형태의 TimeSeries 환경 데이터
(datetime, supplyEC, supplyPH, innerCO2, innerHum, innerTemp, innerSolar)
- csv 형태의 초기 생육 정보
(datetime, 초장, 엽장, 엽폭, 엽병장, 엽수, 관부직경, 화방꽃수, 착화수)

2-2) Output

- csv 형태의 Input Data 추론 결과에 따른 생육 정보
(초장, 엽장, 엽폭, 엽병장, 엽수, 관부직경, 화방꽃수, 착화수)

- 제출 방식 : 사전테스트 문제공개 시점에 각 팀별로 개별 고지되는 사이트에 파일 제출 (업로드)