

사람의 행동분류

Random Forest 의 하이퍼파라미터 변경에 따른 성능결과 분석

18011547 컴퓨터공학과 최예린

제공된 데이터 셋

19 세-48 세, 총 30 명의 피실험자로투버 galaxy s2 를 이용하여 수집한 센서 데이터이다. 이를 통해 소지자의 행동을 분류할 수 있다. 데이터는 50Hz 주기로 수집되었다. 활동의 종류는 1) 걷기 2) 계단 오르기 3) 계단 내려가기 4) 앉기 5) 서기 6) 눕기 로 구성되었다.

experiment number ID	실험 ID
user number ID	유저 ID
activity number ID	행동 ID
Label start point	행동을 시작하는 지점
Label end porint	행동이 끝나는 지점

Random Forest Classifier - hyperparameter 값에 따른 성능분석

이 레포트에서는 n_estimator 와 min_sample_split, max_depth 의 변경에 따른 성능결과를 분석합니다.

no	parameter	value	score
1	-	default	0.93774
2	n_estimator	15	0.93654
3	n_estimator	20	0.93653
4	n_estimator	30	0.93656
5	n_estimator	50	0.93701
6	min_samples_split	3	0.93773
7	min_samples_split	4	0.93772
8	min_samples_split	5	0.93756
9	max_depth	10	0.92956
10	max_depth	15	0.93003
11	max_depth	20	0.93015
12	max_depth	25	0.93756
13	max_depth	30	0.93774

이 데이터 셋에서는

n_estimator default 값일 때 가장 좋은 성능을 보였다.

min_sample_split default 값인 2 일 때 가장 좋은 성능을 보였으며, 값이 커질수록 성능은 좋지 않음을 알 수 있었다.

max_depth default 값일 때의 깊이를 모른채로 파라미터 값만 변경했더니 값이 작을수록 성능이 좋지 않게 나왔다. 아마 깊이가 이것보다 더 깊었을 것으로 예상된다.