

[7주차 과제2] DT의 하이퍼파라미터 변경에 따른 성능결과 분석

센서데이터를 이용한 행동분류를 통한 분석

18011480 이진형 | 컴퓨터공학과

문제 분석

문제에서 주어진 데이터는 가속도 센서와 자이로 센서를 통해 측정된 데이터이다. 데이터 셋을 기반으로 WALKING, WALKINGUPSTAIRS, WALKINGDOWNSTAIRS, SITTING, STANDING, LAYING으로 구분된 여섯 가지의 사람 활동으로 분류하는 것이 목적이다. 당 데이터를 이용, 랜덤 포레스트를 통해 모델 학습을 진행했다.

실험 조건

랜덤 포레스트는 각 트리는 데이터의 일부에 오버피팅하는 경향을 가진다는 데 기초한다. 랜덤 포레스트 모델에서의 주요 파라미터는 생성할 트리의 개수인 `n_estimators`와 선택할 최대 특성의 수인 `max_features`이다. 따라서 이 두가지 파라미터의 변화를 통해 실험을 진행했다.

실험

진행한 실험의 조건과 결과는 다음과 같다.

n_estimators	max_features	Score
1000	auto	0.93581
1000	sqrt	0.93581
1000	log2	0.94062
100	auto	0.93324
100	sqrt	0.93324
100	log2	0.93549
10	auto	0.92522
10	sqrt	0.92522
10	log2	0.89730
baseline(randomforest)		0.99611
baseline(decisiontree)		0.87387

실험 결과

위와 같은 조건으로 실험을 진행한 결과, 최고 점수인 0.94062를 기록한 파라미터는 다음과 같았다.

1000/log2

이번 실험에서는 실험 진행 전에 예상했던 바와 같이 트리를 더욱 많이 생성할 수록 높은 점수가 나온 것을 알 수 있었다.