Dacon 자율주행 센서의 안테나 성능 예측 AI 경진대회

[배경]

안녕하세요 여러분! 🙌 자율주행 센서의 안테나 성능 예측 AI 경진대회에 오신 것을 환영합니다.

Radar는 자율주행 차에 있어 차량과의 거리, 상대 속도, 방향 등을 측정해주는 필수적인 센서 부품입니다.

전기자동차, 자율주행차, 로보택시, 자율주행 택배로봇 등 Radar가 활용되는 시장은 점차 커지고 있으며,

제품 종류도 기존 단거리, 중거리 및 장거리 Radar 뿐 아니라 차량 실내용 및 4D 이미징 Radar 등 다변화 되는 추세입니다.

LG에서는 제품의 성능 평가 공정에서 양품과 불량을 선별하고 있습니다.

그리고 AI 기술을 활용하여 공정 데이터와 제품 성능간 상관 분석을 통해 제품의 불량을 예측/분석하고,

수율을 극대화하여 불량으로 인한 제품 폐기 비용을 감축 시키기 위해 노력하고 있습니다.

[목표]

공정 데이터를 활용하여 Radar 센서의 안테나 성능 예측을 위한 AI 모델 개발

[대회 기간]

• 2022년 8월 01일 10:00 ~ 2022년 8월 26일 17:00

[참가 대상]

- 학생, 일반인 누구나 참가 가능
- (LG Aimers 수강생은 참가 신청 시 LG Aimers 아이디 필수 기재)

[주최/주관]

- 주최 : LG AT Research
- 주관 : 데이콘

자율주행 센서의 안테나 성능 예측 AI 경진대회

알고리즘 | 정형 | 회귀 | 자율주행 | NRMSE

- ₩ 상금:총 1,000만원
- © 2022.08.01 ~ 2022.08.26 16:59

+ Google Calendar

🔐 1,758명 📋 마감





Contents

01. 경진대회 소개

02. 아키텍쳐 개요

03. 결과 정리

01 경진대회 소개



Feature type: Structured Data

Task type: Regression (14 Targets)

전기자동차, 자율주행차, 로보택시, 자율주행 택배로봇 등 다양한 분야에서 Radar의 활용도가 높아지고 가운데, LG에서는 이러한 Radar의 수율을 개선하고자 한다.

또한 AI 기술을 활용하여 공정 데이터와 제품 성능간 상관 분석을 통해 제품의 불량을 예측/분석하고, 수율을 극대화하여 불량으로 인한 제품 폐기 비용을 감축 시키고자 한다.



LG측에서 제공하는 각 Target 별 NRMSE의 가중합

• 평가 산식: Normalized RMSE (NRMSE)

```
def lg_nrmse(gt, preds):
# 각 Y Feature별 NRMSE 총합
# Y_01 ~ Y_08 까지 20% 가중치 부여
all_nrmse = []
for idx in range(1,15): # ignore 'ID'
    rmse = metrics.mean_squared_error(gt[:,idx], preds[:,idx], squared=False)
    nrmse = rmse/np.mean(np.abs(gt[:,idx]))
    all_nrmse.append(nrmse)
    score = 1.2 * np.sum(all_nrmse[:8]) + 1.0 * np.sum(all_nrmse[8:14])
    return score
```

Public score : 전체 테스트 데이터 중 30%
Private score : 전체 테스트 데이터 중 70%

아키텍쳐 개요



✓ Top 3 Leaderboard Score를 기록한 모델들의 average ensemble로 최종 예측치를 산출

1. H2O AutoML + Vanilla Feature set

LB Score: 1.93776

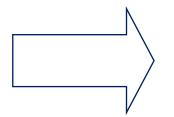
2. H2O AutoML + Feature Engineering

LB Score: 1.94184

3. H2O AutoML + Feature Engineering

LB Score: 1.94433

Average Ensemble



LB Score: 1.93610

Final Score :1.9594 (Top 6%)

02

아키텍쳐 개요 – Feature type

	Feature		설명	타입
0	X_01	PCB 체결 시	단계별 누름량(Step 1)	numeric
1	X_02	PCB 체결 시	단계별 누름량(Step 2)	numeric
2	X_03		방열 재료 1 무게	numeric
3	X_04		1차 검사 통과 여부	binary
4	X_05	PCB 체결 시	단계별 누름량(Step 3)	numeric
5	X_06	PCB 체결 시	단계별 누름량(Step 4)	numeric
6	X_07		방열 재료 1 면적	numeric
7	X_08		방열 재료 2 면적	numeric
8	X_09		방열 재료 3 면적	numeric
9	X_10		방열 재료 2 무게	numeric
10	X_11		방열 재료 3 무게	numeric
11	X_12		커넥터 위치 기준 좌표	numeric
12	X_13	각 안테	나 패드 위치(높이) 차이	numeric
13	X_14		1번 안테나 패드 위치	numeric
14	X_15		2번 안테나 패드 위치	numeric
15	X_16		3번 안테나 패드 위치	numeric
16	X_17		4번 안테나 패드 위치	numeric
17	X_18		5번 안테나 패드 위치	numeric

18	X_19	1번 스크류 삽입 깊이	numeric
19	X_20	2번 스크류 삽입 깊이	numeric
20	X_21	3번 스크류 삽입 깊이	numeric
21	X_22	4번 스크류 삽입 깊이	numeric
22	X_23	2차 검사 통과 여부	binary
23	X_24	커넥터 1번 핀 치수	numeric
24	X_25	커넥터 2번 핀 치수	numeric
25	X_26	커넥터 3번 핀 치수	numeric
26	X_27	커넥터 4번 핀 치수	numeric
27	X_28	커넥터 5번 핀 치수	numeric
28	X_29	커넥터 6번 핀 치수	numeric
29	X_30	스크류 삽입 깊이1	numeric
30	X_31	스크류 삽입 깊이2	numeric
31	X_32	스크류 삽입 깊이3	numeric
32	X_33	스크류 삽입 깊이4	numeric
33	X_34	스크류 체결 시 분당 회전수 1	numeric
34	X_35	스크류 체결 시 분당 회전수 2	numeric
35	X_36	스크류 체결 시 분당 회전수 3	numeric
36	X_37	스크류 체결 시 분당 회전수 4	numeric
37	X_38	하우징 PCB 안착부 1 치수	numeric

38	X_39	하우징 PCB 안착부 2 치수	numeric
39	X_40	하우징 PCB 안착부 3 치수	numeric
40	X_41	레이돔 치수 (안테나 1번 부위)	numeric
41	X_42	레이돔 치수 (안테나 2번 부위)	numeric
42	X_43	레이돔 치수 (안테나 3번 부위)	numeric
43	X_44	레이돔 치수 (안테나 4번 부위)	numeric
44	X_45	안테나 부분 레이돔 기울기	numeric
45	X_46	실란트 본드 소요량	numeric
46	X_47	3차 검사 통과 여부	binary
47	X_48	4차 검사 통과 여부	binary
48	X_49	Cal 투입 전 대기 시간	numeric
49	X_50	RF1 부분 SMT 납 량	numeric
50	X_51	RF2 부분 SMT 납 량	numeric
51	X_52	RF3 부분 SMT 납 량	numeric
52	X_53	RF4 부분 SMT 납 량	numeric
53	X_54	RF5 부분 SMT 납 량	numeric
54	X_55	RF6 부분 SMT 납 량	numeric
55	X_56	RF7 부분 SMT 납 량	numeric

아키텍쳐 개요 – Preprocessing & Feature Engineering

Preprocessing

1. 필요 작업 없음

√ Feature Engineering sets

- V1. Vanilla
- V2. Drop features(var=0)
- V3. Drop features(var=0) + Sequence features(n=3)
- V4. Drop features(var=0) + Polynomial features(n=2) + Interaction features(n=2)
- V5. Drop features(var=0) + Mean on related-features + Standard deviation on related-features
- V6. Drop features(var=0) + Polynomial features(n=2) + Interaction features(n=2)) + Mean on related-features + Standard deviation on related-features
- V7. Drop features(var=0) + Sequence feature + Polynomial features(n=2) + Interaction features(n=2)) + Mean on related-features + Standard deviation on related-features

02

아키텍쳐 개요 – Preprocessing & Feature Engineering

	Feature			설명	타입
0	X_01	PCB 체결	시 단계별	별 누름량(Step 1)	numeric
1	X_02	PCB 체결	시 단계별	별 누름량(Step 2)	numeric
2	X_03			방열 재료 1 무게	numeric
3	X_04		1	차 검사 통과 여부	binary
4	X_05	PCB 체결	시 단계별	별 누름량(Step 3)	numeric
5	X_06	PCB 체결	시 단계별	별 누름량(Step 4)	numeric
6	X_07			방열 재료 1 면적	numeric
7	X_08			방열 재료 2 면적	numeric
8	X_09			방열 재료 3 면적	numeric
9	X_10			방열 재료 2 무게	numeric
10	X_11			방열 재료 3 무게	numeric
11	X_12		커넥	터 위치 기준 좌표	numeric
12	X_13	각 안	테나 패드	위치(높이) 차이	numeric
13	X_14		1번	안테나 패드 위치	numeric
14	X_15		2번	안테나 패드 위치	numeric
15	X_16		3번	안테나 패드 위치	numeric
16	X_17		4번	안테나 패드 위치	numeric
17	X_18		5번	안테나 패드 위치	numeric

18	X_19	1번 스크류 삽입 깊이	numeric
19	X_20	2번 스크류 삽입 깊이	numeric
20	X_21	3번 스크류 삽입 깊이	numeric
21	X_22	4번 스크류 삽입 깊이	numeric
22	X_23	2차 검사 통과 여부	binary
23	X_24	커넥터 1번 핀 치수	numeric
24	X_25	커넥터 2번 핀 치수	numeric
25	X_26	커넥터 3번 핀 치수	numeric
26	X_27	커넥터 4번 핀 치수	numeric
27	X_28	커넥터 5번 핀 치수	numeric
28	X_29	커넥터 6번 핀 치수	numeric
29	X_30	스크류 삽입 깊이1	numeric
30	X_31	스크류 삽입 깊이2	numeric
31	X_32	스크류 삽입 깊이3	numeric
32	X_33	스크류 삽입 깊이4	numeric
33	X_34	스크류 체결 시 분당 회전수 1	numeric
34	X_35	스크류 체결 시 분당 회전수 2	numeric
35	X_36	스크류 체결 시 분당 회전수 3	numeric
36	X_37	스크류 체결 시 분당 회전수 4	numeric
37	X_38	하우징 PCB 안착부 1 치수	numeric

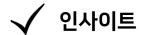
38	X_39	하우징 PCB 안착부 2 치수 numeric
39	X_40	하우징 PCB 안착부 3 치수 numeric
40	X_41	레이돔 치수 (안테나 1번 부위) numeric
41	X_42	레이돔 치수 (안테나 2번 부위) numeric
42	X_43	레이돔 치수 (안테나 3번 부위) numeric
43	X_44	레이돔 치수 (안테나 4번 부위) numeric
44	X_45	안테나 부분 레이돔 기울기 numeric
45	X_46	실란트 본드 소요량 numeric
46	X_47	3차 검사 통과 여부 binary
47	V /0	4차 검사 통과 여부 binary
47	X_48	4차 검사 통과 여부 binary
48	X_48 X_49	Cal 투입 전 대기 시간 numeric
48	X_49	Cal 투입 전 대기 시간 numeric
48 49	X_49 X_50	Cal 투입 전 대기 시간 numeric RF1 부분 SMT 납 량 numeric
48 49 50	X_49 X_50 X_51	Cal 투입 전 대기 시간 numeric RF1 부분 SMT 납 량 numeric RF2 부분 SMT 납 량 numeric
48 49 50 51	X_49 X_50 X_51 X_52	Cal 투입 전 대기 시간 numeric RF1 부분 SMT 납 량 numeric RF2 부분 SMT 납 량 numeric RF3 부분 SMT 납 량 numeric
48 49 50 51 52	X_49 X_50 X_51 X_52 X_53	Cal 투입 전 대기 시간 numeric RF1 부분 SMT 납 량 numeric RF2 부분 SMT 납 량 numeric RF3 부분 SMT 납 량 numeric RF4 부분 SMT 납 량 numeric

아키텍쳐 개요 - Model (H2O AutoML)

Training & Inference - H2O

```
1 h2o.init(nthreads=-1, max_mem_size=8)
 3 seed_everything()
 4 for cnt, value in enumerate(allTarget):
      tmp_df = h2o.H20Frame(pd.concat([df_full_x.astype("float32"), df_full_y[[value]].astype("float32")], axis=1))
      feature_names = list(df_full_x.columns)
      target_name = value
      aml = h2o.automl.H2OAutoML(max_runtime_secs=int(3600 * 24 / len(allTarget)), seed=42, project_name="LG_selfDriv
10
                                 nfolds=5, stopping_metric="MAE", stopping_rounds=100)
11
      aml.train(training_frame=tmp_df, x=feature_names, y=target_name)
12
      print(aml.leaderboard.head())
13
      inference_root_path = folder_path + "inference/" + "featureV" + str(feature_version) + "+" + "H2OAutoML" + "/"
14
      createFolder(inference_root_path)
15
      if os.path.exists(inference_root_path + "submission_" + "featureV" + str(feature_version) + "+" + "H2OAutoML"
16
          submission = pd.read_csv(inference_root_path + "submission_" + "featureV" + str(feature_version) + "+" + "\
17
      else:
18
          submission = pd.read_csv(folder_path + 'rawdata/sample_submission.csv')
19
      submission[value] = aml.predict(h2o.H20Frame(df_test_x.astype("float32"))).as_data_frame().to_numpy().flatten()
20
      submission.to_csv(inference_root_path + "submission_" + "featureV" + str(feature_version) + "+" + "H20AutoML"
21
22 h2o.cluster().shutdown()
```

03 결과 정리



- 1. 관련된 feature들에 대한 표준편차를 파생변수로 추가 할 시 성능 향상이 있었다.
- 이는 즉, 관련된 feature들 간의 값 편차와 target간의 상관성이 있다는 것을 알 수 있다. (LB Score 기준 약 0.003 의 NRMSE 감소를 보임)
- 2. 대부분의 정형데이터의 특성과 같이 DL 모델은 다소 성능이 떨어지는 결과를 보였다.
- 3. EDA 및 상관분석 결과 0.5 이상의 양 또는 음의 상관성을 보이는 feature가 없었다.
- 4. 주기성을 보이는 feature도 발견을 하였으나, Domain knowledge의 부재로 원인을 파악할 수 없었다.

감사합니다