

자동차용 침입 탐지 트랙

데이터셋 설명

2019. 10. 1.

목차

1	챌린지 문제	2
2	차량 내부 네트워크 데이터셋	2
2.1	기본정보.....	2
2.1.1	파일 정보.....	2
2.1.2	학습용 데이터셋 - CAN data	2
2.2	제출용 데이터셋	3

1 챌린지 문제

주어진 차량 내부 네트워크 데이터셋을 분석하여, 차량내 침입 탐지를 할 수 있는 알고리즘을 제안하는 것이 본 트랙의 주요 목표입니다. 알고리즘은 룰, 머신러닝, AI 기반 등 형태에 제한은 없으나, 프로그램으로 구현이 가능해야 합니다. (즉, 자동화가 가능해야 합니다.)

2 차량 내부 네트워크 데이터셋

2.1 기본정보

2.1.1 파일 정보

데이터셋은 학습용과 제출용으로 나누어져 있으며, 학습용 데이터셋의 경우 차량별 4개씩 전체 12개의 파일로 구성됩니다. 제출용 데이터셋은 레이블링 (Class)을 제외한 내용으로 이뤄지며, 총 9개의 파일로 구성됩니다. 압축 해제 시 학습용 데이터셋 약 60MB, 제출용 데이터셋 24MB 입니다.

정상	CHEVORLET_SPARK_Attack_free_TRAIN_Release.csv	SPARK CAN Dataset
Flooding 공격	CHEVORLET_SPARK_Flooding_TRAIN_Release.csv	
Fuzzy 공격	CHEVORLET_SPARK_Fuzzy_TRAIN_Release.csv	
Malfunction 공격	CHEVORLET_SPARK_Malfunction_TRAIN_Release.csv	
정상	HYUNDAI_SONATA_Attack_free_TRAIN_Release.csv	SONATA CAN Dataset
Flooding 공격	HYUNDAI_SONATA_Flooding_TRAIN_Release.csv	
Fuzzy 공격	HYUNDAI_SONATA_Fuzzy_TRAIN_Release.csv	
Malfunction 공격	HYUNDAI_SONATA_Malfunction_TRAIN_Release.csv	
정상	KIA_SOUL_Attack_free_TRAIN_Release.csv	SOUL CAN Dataset
Flooding 공격	KIA_SOUL_Flooding_TRAIN_Release.csv	
Fuzzy 공격	KIA_SOUL_Fuzzy_TRAIN_Release.csv	
Malfunction 공격	KIA_SOUL_Malfunction_TRAIN_Release.csv	

Figure 1. 학습용 데이터셋 파일

File_1.csv
File_2.csv
File_3.csv
File_4.csv
File_5.csv
File_6.csv
File_7.csv
File_8.csv
File_9.csv

Figure 2. 제출용 데이터셋 파일

2.1.2 학습용 데이터셋 - CAN data

Timestamp	CAN ID	DLC	Data field	Class
1513921661.084060	02C0	8	14 00 00 00 00 00 00 00	R
1513921661.084294	0329	8	0C B9 7F 14 11 20 00 14	R
1513921661.084535	0545	8	D8 00 00 8C 00 00 00 00	R

Figure 3. CAN Data – 학습용

Timestamp : CAN Data 덤프 시, 해당 CAN packet이 로깅되는 시간

CAN ID : CAN packet의 식별 ID

DLC : CAN packet의 Data field의 크기

Data field : CAN packet의 데이터가 적재된 필드

Class : CAN Data 추출 시, 정상 (R) 또는 비정상 (T) 를 나타내기 위한 필드

2.2 제출용 데이터셋

Number	Timestamp	CAN ID	DLC	Data field
1,	1513920124.163693,	019D,	8,	43 2C 3C D3 00 0C 81 7F
2,	1513920124.163929,	018E,	8,	00 00 00 69 B6 9A 06 9A
3,	1513920124.164150,	01A1,	7,	00 00 01 41 6A 6A 00

Figure 4. CAN Data - 제출용

학습용/제출용 데이터셋의 차이는 각 데이터셋에 Class 필드의 유무입니다. 제출용 데이터셋에는 학습용과 다르게, Class 필드가 삭제되어 있으며, Timestamp 필드 앞에 각 라인의 번호를 의미하는 'Number' 필드가 추가되어 있습니다. 제출용 데이터셋에 대한 평가를 내린 후, 제출용 데이터마다 'Number'와 'Class'를 추가하여 제출하시면 됩니다.

Number	Class
1,	R
2,	R
3,	T
4,	T
5,	R
6,	T

Figure 5. 제출용 데이터셋에 대한 판별 파일

Figure 5와 같이, 제출용 데이터셋에 대한 판별이 끝난 이후, 각 파일의 Number, Class 의 형식으로 파일의 내용을 작성해주시면 됩니다.