

비지도 앙상블 군집화를 통한 차와 차 사이의 충돌 시나리오 분류

신승륜, 이청화, 옥기수, 홍성수*
*서울대학교

{srshin, chlee, ksok, *sshong}@redwood.snu.ac.kr

Vehicle-to-Vehicle Crash Scenario Classification through Unsupervised Ensemble Clustering

Seungryun Shin, Cheonghwa Lee, Kisu Ok, Seongsoo Hong*
*Seoul National University

요 약

시장에 출시되는 모든 자율주행차는 안전성 평가를 거쳐야 한다. 안전성 평가는 일반적으로 시뮬레이션이나 PG(Proving Ground)에서 사전 테스트를 거친 후, 실제 도로에서 주행하면서 이루어진다. 사전 테스트를 통해 실제 주행 테스트에서는 많은 비용과 시간을 절약할 수 있고 사고가 날 위험성을 줄일 수 있다. 하지만 실 주행 테스트에서 사전 테스트에서 고려하지 못한 돌발 및 위험 상황이 발생할 수가 있다. 이러한 문제를 해결할 방법 중 하나는 자율주행차 및 엡지 인프라에서 수집한 실제 사고 데이터를 수집 및 분석하여 돌발 및 위험 상황의 패턴을 찾고, 이를 빈도와 정도에 따라 분류/군집하여 최대한 많은 위험 상황을 사전 테스트에서 수행하는 것이다. 본 연구에서는 차와 차, 차와 사람, 차와 환경 사이 등 다양한 사고 종류 중에서 차와 차 사이의 사고 상황을 다룬다. 차와 차 사이의 사고 상황을 담고 있는 데이터는 data-driven 기계학습 방식 중 K-Means 비지도학습 방식의 Extreme Gradient Boosting (XGBoost) 알고리즘을 사용하여 군집 및 분류한다. 해당 방식은 여러 기계학습 모델을 사용하여 1 차 결과를 출력하고, 이를 취합하여 최종 결과를 내는 ensemble 방식이다. 이 방식은 하나의 모델을 사용하는 방식에 비해 분류의 일반화 및 높은 정확도를 얻을 수 있다. 뿐만 아니라 전체 프로세스가 기계학습이므로 자동화가 가능하다.

I. 서론

자율주행차의 안전성을 평가하는 방법은 주행기반 검증 방법과 시나리오 기반의 방법으로 나뉜다. 주행기반 검증 방법은 실 도로에서 수행하기 때문에 이상적(ideal)인 방법이지만 실제로 수행하기에는 돈과 시간이 들 뿐만 아니라 검증 중에 주변 차량에게 위험을 야기하는 등의 심각한 애로사항들이 있다 [1]. 최근에는 자율주행차의 주행 중에 발생할 수 있는 위험 상황을 하나의 검증 시나리오를 만들어 미시적인 관점에서 하나하나 검증하고, 이를 취합하여 거시적인 관점에서 안정성을 평가하는 시나리오 기반 검증 방법을 연구하는 추세다 [2].

시나리오 기반 검증 방법에서는 적절한 위험 시나리오를 생성하는 것이 중요하다. 이는 사람이 도메인 지식을 사용하는 지식(knowledge) 기반의 방식과 주행 데이터에서 패턴을 찾아 사용하는 data-driven 방식으로 나뉜다 [3]. 최근 기계학습의 실용적인 발전으로 인해 data-driven 방식으로 시나리오를 만드는 연구가 대세를 이루고 있다. 주행 데이터는 자율주행차에서 수집할 뿐만 아니라 최근에는 엡지 인프라에서 다양한 데이터를 수집하고 있다. 이러한 데이터 속에서 기계학습의 비지도 방식으로 특징 추출을 통해 위험 상황에 대한 패턴을 찾고, 군집화를 통해 위험 상황들을 분류한다. 군집화를 통해 unlabeled class 의 시나리오 데이터를 pseudo-

labeled class 가 부여된 시나리오 데이터로 만드는 것이다 [4].

기존의 연구들에서는 이러한 차량 데이터를 군집화 하는데 있어서 Random Forest Classifier (RFC) 와 같은 single 모델을 사용하는 방식을 주로 썼다. 본 논문에서는 multi 모델을 사용하는 ensemble 기법으로 차량 데이터를 군집화 및 분류할 것을 제안한다. 우리가 사용하는 대표적인 ensemble 기법 중 하나인 XGBoost 는 여러 single 모델들에서 출력된 값을 취합할 때 Bagging 과 Boosting 을 혼합한 방식을 사용한다. Bagging 은 multi 모델들 간에 병렬 구조를 사용하고, Boosting 은 직렬/순차 구조를 사용한다 [5]. 이런 ensemble 기법은 RFC 와 같은 Single 알고리즘 방식에 비해 차와 차 사이의 사고 상황 분류에서 일반화 및 높은 정확도를 가질 것으로 기대된다.

II. 본론

본 연구는 총 세 단계로, (1) 차와 차 사이의 사고 dataset 선정, (2) 비지도 충돌 사고 상황 군집화(clustering), (3) 알고리즘 방식의 평가 및 비교(evaluation) 나뉜다. 첫 번째 단계에서는 차량 대 차량 사고와 관련된 오픈 소스 주행 데이터를 선택한다. 두 번째 단계에서는 K-Means clustering 알고리즘을 사용하여 위험한 상황을 포함한 주행 데이터를 군집화한다. 구체적으로 RFC 와 XGBoost 를 통해 주행

데이터에서 차량 대 차량 충돌 사고 상황만 검출하고, 이를 다시 전방 충돌, 측방 충돌, 후방 충돌 그룹화하는 모델을 학습시킨다. 세 번째 단계에서는 최종 출력 정확도와 모델의 일반화를 평가하고 비교한다.

1. 차대차 사고 Dataset 선정

통신과 컴퓨터의 발달로 많은 차량 주행 데이터가 오픈 소스 형식으로 제공되고 있다. 그 중 본 논문에서는 독일 국토에서 차량의 움직임을 드론으로 관찰 및 측정하는 RWTH Aachen University의 Highway Drone (highD) dataset를 실험에서 사용한다. HighD dataset은 독일의 6개 국토에서 드론을 통해 일정 기간 하늘에서 110,500대 이상의 차량을 측정했다. 이 dataset은 10cm 미만의 정확도를 가지고 있으며 수많은 교통 상황과 차량 운전 방법을 포함하고 있다. 데이터 유형에는 총 348,750개의 이벤트와 25개의 기능이 있다.

2. Clustering (군집화)

군집화 방법은 그림 1에 묘사되어 있다. 먼저, K-Means 알고리즘으로 자동차 사고 시 생성되는 유사한 수치 데이터를 기능별로 clustering 하여 label을 부여한다. K-Means 알고리즘의 경우 클러스터의 수를 k 라는 하이퍼 파라미터로 설정해야 한다. 클러스터의 수 k 는 Elbow Curve Method를 사용하여 결정하는데, Elbow Curve Method는 임의의 k 숫자들을 정해 알고리즘을 실행했을 때 값이 급격히 떨어지는 부분의 k 값을 이용하여 모든 데이터 포인트의 중심점으로부터의 거리의 평균을 구하는 방법이다.

이렇게 label이 부여된 주행 데이터를 대상으로 RFC와 XGBoost를 통해 차와 차 사이의 사고 상황들을 분류한다. RFC는 여러 기계학습 모델을 동시에 생성 및 사용하여 1차적으로 분류 결과를 출력하고, 이를 집계하여 최종 결과를 도출한다. XGBoost는 RFC와는 다르게 각 기계학습 모델들이 이전 모델의 결과를 기반으로 하이퍼 파라미터를 업데이트하여 순차적으로 모델을 생성한다. 이는 각 모델별로 분류된 차량 대 차량 사고 상황에 대해 다음에 형성될 모델의 정확도에 영향을 미친다. 그런 다음 RFC와 마찬가지로 각 의사결정 트리의 출력 값을 수집하고 투표하여 가장 많은 표를 얻은 값을 모델의 최종 값으로 선택한다. 이렇게 하여 형성된 군집들은 차와 차 사이의 사고 상황들인 전방 충돌, 측방 충돌, 후방 충돌로 분류한다.

3. Evaluation (모델 평가)

마지막으로 RFC와 XGBoost의 최종 출력 정확도와 일반화를 비교 검증한다. XGBoost의 출력 정확도와 RFC 정확도는 각각적으로 비교할 수 있으나, 일반화는 learning curve를 통해 확인할 수 있다. learning curve는 모델 학습 오류와 테스트 데이터셋에 관한 관계 그래프다. 일반화가 있는지 확인하려면 learning curve 그래프를 통해 과적합되었는지 본다. 과적합은 모델 훈련 오류가 테스트 데이터 세트 오류보다 작은 경우이다.

III. 결론

본 논문에서는 차량 대 차량 사고 상황 분류의 어려움을 비지도 양상불 학습 알고리즘인 K-Means Clustering과 XGBoost를 통해 해결할 것을 제안했다.

ACKNOWLEDGMENT

본 논문은 정보통신기획평가원 자율주행 기술개발 혁신 사업(과제 번호:2021-0-00697)의 지원을 받아 수행된 연구임 (과제명: 엣지 기반 자율주행 기법의 Fall

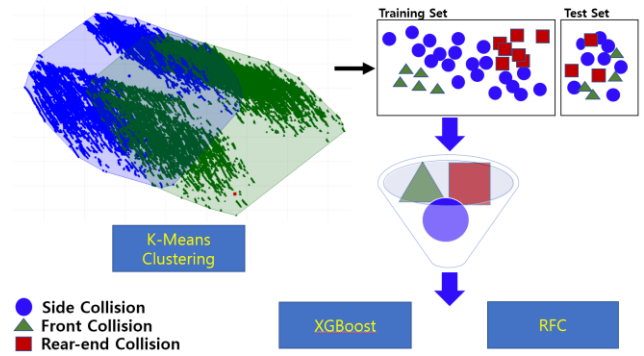


그림 1. 제안된 ensemble 방식인 XGBoost 모델과 single 알고리즘 방식인 RFC 모델을 이용한 데이터 군집화

back MRC에 따른 운영권 SW 안전성 및 대응방안 검증 기술 개발)

참 고 문 헌

- [1] Tang, S., "A survey on Automated Driving System Testing: Landscapes and trends," arXiv.org, Jun. 2022, (<https://arxiv.org/abs/2206.05961>).
- [2] King, C., "A taxonomy and survey on validation approaches for Automated Driving Systems," IEEE Xplore, 2020, (<https://ieeexplore.ieee.org/document/9272219>).
- [3] Nalic, D., "Scenario based testing of Automated Driving Systems: A literature survey," FISITA Web Congress, 2020, (https://www.researchgate.net/profile/Demin-Nalic-2/publication/344896382_Scenario_Based_Testing_of_Automated_Driving_Systems_A_Literature_Survey/links/600806cba6fdccdb868bf72/Scenario-Based-Testing-of-Automated-Driving-Systems-A-Literature-Survey.pdf?origin=publication_detail).
- [4] Kruber, F., "Unsupervised and supervised learning with the random forest algorithm for traffic scenario clustering and classification," IEEE Xplore, 2019, (<https://ieeexplore.ieee.org/document/8813994>).
- [5] Chen, T., "XGBoost: A Scalable Tree Boosting System," ACM Conferences, Aug. 2016, (<https://dl.acm.org/doi/10.1145/2939672.2939785>).