

# 고령 운전자의 안전 운전을 위한 위험 상황 시나리오 구축 -고속도로 교통사고 기반의 토픽모델링을 활용하여-

채오성, 최낙현, 김정화

경기대학교

yrias@gmail.com., [nadosan1@kyonggi.ac.kr](mailto:nadosan1@kyonggi.ac.kr), [junghwa.kim@kyonggi.ac.kr](mailto:junghwa.kim@kyonggi.ac.kr)

## A Study on the Scenario development for Old driver with Topic Modeling based on Highway Crash Data

Chae Oh Seong, Choi Nak Hyeon, Kim Jung Hwa

Kyonggi Univ

### 요약

본 논문은 고령화로 인한 다양한 문제 중, 도로 위 교통사고는 치명적인 인명사고로 이어질 수 있으므로 충분한 대처와 사고율 감소를 위한 전략이 필요하지만, 고령 운전자 교통사고는 꾸준히 그 수를 유지하고 있다. 고령 운전자의 운전 능력을 일반 운전자의 운전 능력과 비교해보았을 때, 고령 운전자의 주사고 발생 원인은 신체의 노화로 인한 운전 능력 저하에서 비롯된다고 할 수 있다. 고령 운전자의 운전 특성을 분석한 여러 연구들이 존재하지만, 사고원인이나 사고 발생률이 높은 이유에 대해 포괄적인 이유가 신체적 노화로 인해 발생한다는 사실만 알 수 있으며, 어떤 특정한 상황에서 고령 운전자 사고가 자주 발생하는지는 알 수 없다. 이에 본 연구에서는 고령 운전자의 안전 운전을 위한 시나리오를 토픽모델링을 통해 구축하였다. 최종적으로 육하원칙에 기반한 키워드 조합을 통해 고령 운전자의 교통사고 위험 상황 시나리오를 총 4가지 유형의 19가지 시나리오를 제시하였다.

### I. 서론

현대의 고령화 문제는 사회문제를 다루는데 있어 무시할 수 없는 큰 부분으로 자리 잡고 있다. 국제기구를 포함한 많은 나라에서 인구를 나이에 따라 구분하는데 UN의 기준에 따르면 만 65세 이상의 인구를 고령인구라 부르며, 고령인구가 전체인구 대비 14% 이상이라면 고령화 사회 20% 이상이라면 초고령 사회로 구분한다. 현재 우리나라의 고령화 비율은 17.1%로 16년 대비 3% 이상 증가한 것으로 보아 초고령화 사회로 향해가고 있음을 짐작할 수 있다(통계청, 2021). 고령화로 인한 다양한 문제 중, 도로 위 교통사고는 치명적인 인명사고로 이어질 수 있으므로 충분한 대처와 사고율 감소를 위한 전략이 필요하지만, TAAS 교통사고 분석시스템의 고령 운전자 사고 데이터를 살펴보면 2016년도 24,429건에서 2020년 31,072건으로 매년 점진적으로 증가하는 것을 보아 고령 운전자에 대한 파악이 미흡한 것을 알 수 있다. 고령 운전자의 경우가 상황 판단에 대한 정답률이 낮고 반응시간이 더 길게 나타났다(정미경 외 1인, 2019). 또한 고령 운전자의 주사고 발생 원인은 대부분이 전방 주시 태만이고, 그 외는 인적요인이 없어 사고 상황을 정확하게 판단하기 힘든 사고가 대부분이다(오주석 외 3인, 2015). 고령 운전자의 운전 특성을 분석한 기존 연구들을 통해 사고원인이나 사고 발생률이 높은 이유에 대한 포괄적인 이유가 신체적 노화로 인한 발생에 대한 분석은 부족한 실정이다. 이에 본 연구는 실제 교통사고 데이터를 통해 고령 운전자의 사고를 유형화하여 분류하고 많이 발생하는 사고 케이스에 대한 원인을 분석하고자 한다. 최종적으로, 분석 데이터를 통해 고령 운전자의 안전 운전을 위한 위험 상황 시나리오를 도출하여 제시하고자 한다.

### II. 본론

본 연구에서는 고령 운전자들의 안전 운전을 위해 고속도로에서

자주 발생하는 사고 유형을 파악하고 이러한 사고를 대비하기 위해 실제 도로 상황이 반영된 고령 운전자의 대표 위험 시나리오를 설계하였다. 시나리오의 실제 도로 상황을 고려하기 위해 2016~2021년 총 6년간의 10,135건 한국의 고속도로 교통사고 정황을 설명한 사고 조사 데이터를 사용하였다. 데이터의 경우 비정형 데이터로 이루어져 있으며, 사고 날짜, 도로 선형, 도로 경사, 절·성토 구분, 발생지점, 교통상황 장애 요인, 도로 환경, 주 사고원인, 사고유형, 사고 직전 차량 조작, 주 사고유형, 주·야, 날씨, 사고상황 등을 포함한다.

고령 운전자 위험 상황 시나리오를 도출하기에 앞서 사고 특성을 분석하기 위해 10,135건의 고속도로 사고 데이터 중 이상치 1,361건을 제거한 8,774건을 사용하여 만 65세 이상 운전자사고 DB를 구축하였다. 구축된 고령 운전자사고 DB를 살펴보면, 고령 운전자 사고는 805건으로 전체 사고의 약 10%를 차지하는 것으로 나타났다. 해당 고령 운전자 사고 데이터는 비정형 데이터들로 구축되어 있어 데이터를 분석하기에 있어 분석자의 주관적 견해가 포함될 가능성이 높고, 유의미한 비정형 데이터와 무의미한 비정형 데이터를 분류하기가 어렵다. 뿐만 아니라, 서로 다른 비정형 데이터간의 관계 또한 분석이 어렵기에 본 연구에서는 이러한 한계점들을 극복하고자 토픽모델링을 사용하여 고령 운전자 사고 DB를 분석하였다. 토픽 그룹들을 해석할 시 주관적 견해를 최대한 배제하기 위해 사고 데이터를 Pegasus Layer를 활용하여 비슷한 유형의 단어계열로 그룹핑하여 분류하였다.

Topic Modeling을 통해 위험 상황 시나리오를 위한 키워드를

추출하였다. 토픽모델링 결과에서는 각 layer 별 Topic의 순서로 Topic의 중요도를 설명한다. 고령 운전자 사고 DB의 경우, 토픽의 수가 4개일 때까지 Coherence score가 증가하다 이후 감소와 증가를 반복하는 형태를 나타 냈지만, 그 영향력이 크게 증가하지 않았으므로, 토픽을 분석할 시 그룹 수를 4개로 설정하는 것이 가장 적합하다 판단하였다.

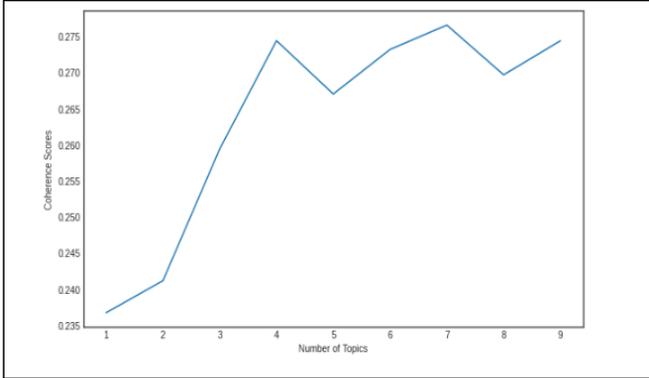


그림 1. 고령 운전자 사고 DB의 Coherence Score

G1 사고상황의 1번 그룹에 해당하는 토픽을 살펴보면, Layer 2 와 Layer 4의 요소만을 포함하고 있는 것을 알 수 있다. Layer 2-도로시설 및 규칙(영구)의 요소로는 차로, 시설물, 도로 방향이 해당한다. Layer 4-도로 주제 행동의 요소로는 차량 행동, 사고 차량 위치, 운전자, 사고 발생 횟수, 추돌 부위, 사고원인이 해당한다. 각 토픽의 비율을 살펴보면, 평균(52.03%) 보다 큰 요소는 차로의 경우, 2차로, 갓길을 제외한 모든 차로가 해당한다. 시설물은 가드레일, 차량 행동은 주행 중, 사고 차량 위치는 12시 방향, 운전자는 피해자 차량, 사고 발생 횟수와 추돌 부위는

Layer	토픽	횟수	전체	비율	
Layer 2	차로	1차로	98	175	56.0%
		2차로	76	161	47.2%
		3차로	25	40	62.5%
		4차로	15	24	62.5%
		갓길	78	162	48.1%
	1, 2 차로 걸침	11	18	61.1%	
	시설물	가드레일	22	41	53.7%
	중앙 분리대	18	42	42.9%	
도로방향	부산 방향	12	53	22.6%	
Layer 4	차량 행동	정차	119	239	49.8%
		주행 중	247	438	56.4%
	사고 차량 위치	12시 방향	77	144	52.4%
	운전자	가해 차량	68	129	52.7%
		피해 차량	43	76	56.6%
	사고 발생 횟수	1차 사고	18	31	58.1%
		2차 사고	24	38	63.2%
	추돌 부위	후미 추돌	61	101	60.4%
		전면 추돌	39	67	58.2%
	사고 원인	중심을 잃고	15	48	31.3%
전방 주시 태만		14	37	37.8%	
졸음 운전		13	22	59.1%	

표 1. G1 1번 그룹 토픽

모든 항목이 포함된다. 마지막으로, 사고원인은 졸음운전이 해당한다. 가장 비율이 높은 요소는 사고 발생 횟수 중 2차 사고가 63.2%로 차지했다.

### III. 결론

본 연구에서는 국내 고속도로 교통사고 데이터와 토픽모델링을 활용하여 실제 사고 데이터 기반의 고령 운전자 위험 상황 시나리오를 개발하였다. 수집된 고속도로 사고 데이터를 pegasus layer 기반으로 전처리하여 비정형 데이터 사고상황을 생성하였다. 또한, 생성된 사고상황은 토픽모델링을 통해 추출한 키워드들을 육하원칙을 기준으로 구축하였다. 육하원칙의 경우, 언제(L5) 어디서(L1, L2) 누가(L4) 무엇을(L2, L3, L4) 왜(L4) 어떻게(L4)의 순서로 적용하였다. 토픽 결과를 살펴보면, G1에는 모든 Layer의 영향을 골고루 받기에 G2 ~ G4의 대부분 요소의 영향을 받는 것으로 나타났다. G2는 환경요소의 영향을 거의 받지 않고 도로시설물의 영향을 많이 받는 그룹이나, 시설물 관련 사고 건수가 적은 만큼 시설물 요소는 유효한 토픽에 포함되지 못하였지만, 환경 차이에 상관없이 발생한 사고 토픽들이 분포해 있다. G3는 일시적인 도로 상황보다 환경요소의 영향을 더 많이 받는 그룹으로 주간, 맑은 날 발생한 사고를 기준으로 복합적인 사고가 발생한 그룹이다. 마지막으로 G4는 도로시설물의 영향을 받지 않는 그룹으로 도로선형에 따른 사고상황이 포함된 그룹이다. 해당 그룹들에서 유의미한 키워드들을 도출하여 G1 ~ G4의 그룹에서 총 4가지 유형의 19가지 위험 상황 대표 시나리오를 생성하였다 (표 17). 시나리오를 생성할 시 육하원칙에 따라 생성해야 하므로, Layer 2와 Layer 4의 키워드가 주인 그룹 G에서 뽑아낸 키워드만으로는 Layer가 부족하여 시나리오 생성에 제약이 생긴다. 때문에, 사전에 전체 사고를 토픽모델링하여 만든 그룹 A를 통해서 시나리오의 생성 한계를 극복하고자 하였다.

### ACKNOWLEDGMENT

This research was supported by granted(2022-MOIS41-002) of Citizen-customized Life Safety Technology Development Program funded by Ministry of Interior and Safety(MOIS, Korea)

### 참고 문헌

[1] ARAKAWA, Y., MIYAUCHI, K., TAKADA, K., SHINOHARA, M., & FUJII, M. "Research on relationship between cognitive impairment and driving behavior of starting/stopping for elderly driver." In 6th International Conference on Road and Rail Infrastructure., 2021

[2] PERSSON, Diane. "The elderly driver: deciding when to stop. The Gerontologist", 1993, 33.1: 88-91.

[3] Takahashi, R., Kobayashi, M., Sasaki, T., Yokokawa, Y., Momose, H., & Ohhashi, T. "Driving simulation test for evaluating hazard perception: Elderly driver response characteristics." Transportation Research Part F: Traffic Psychology and Behaviour, 2017, 49, pp. 257-270.