

보안과제(), 일반과제(O) / 공개(O), 비공개()
국토교통기술촉진연구사업 최종평가 보고서

발 간 등 록 번 호

11-1613000-001943-01

R&D / 15CTAP-C078824-02

선제적 교통사고 예방을 위한 교통안전관리용 빅데이터 활용 교통사고 위험요인 분석 기술개발 최종보고서

2016. 09. 22.

주 관 연구 기관 / 한국과학기술원

국 토 교 통 부
국토교통과학기술진흥원

제 출 문

국토교통부장관 귀하

‘선제적 교통사고 예방을 위한 교통안전관리용 빅데이터 활용 교통사고 위험요인 분석 기술 개발’(연구개발 기간 : 2015.08.14 ~ 2016.08.13.)의 과제의 최종보고서 2부를 제출합니다.

2016. 09 . 22 .

주관연구기관명 : 한국과학기술원

(대표자) 강성모



주관연구기관책임자: 교수 윤윤진

국토교통부소관 연구개발사업 운영규정 제38조에 따라 최종보고서 열람에
동의합니다.

보고서 요약서

과제 고유 번호	15CTAP-C07882 4-02	해당 단계 연구 기간	2015.08.14.~ 2016.08.13	단계구분	2/2 (해당단계)/(총단계)
연구사업명	중사업명	국토교통기술촉진연구사업			
	세부사업명	창조형 모형			
연구과제명	대과제명	선제적 교통사고 예방을 위한 교통안전관리용 빅데이터 활용 교통 사고 위험요인 분석 기술 개발			
	세부과제명				
연구책임자	윤윤진	해당단계 참여 연구원 수	총: 8명 내부: 명 외부: 명	해당단계 연구개발비	정부:50,000 천원 기업: 천원 정부 외: 천원 계: 천원
		총 연구기간 참여 연구원 수	총: 8명 내부: 명 외부: 명	총 연구개발비	정부:100,000천원 기업: 천원 정부 외: 천원 계: 천원
연구기관명 및 소속 부서명	한국과학기술원 건설 및 환경공학과			참여기업명	
국제공동연구	상대국명:			상대국 연구기관명:	
위탁연구	연구기관명:			연구책임자:	
최근 10여년간 OECD 회원국 기준 최하위권에 머물고 있는 국내 교통 안전 실태를 개선하기 위하여, 본 연구에서는 선 제적 교통사고 예방을 위한 위험 요인 분석기술을 개발함. 정형 및 비정형의 대용량 교통데이터베이스와 교통사고 이 력 데이터들을 통합하고, 여기에 고급 통계기법 및 데이터 마이닝 (data mining) 기술을 적용한 사고 심각도 결정 모 델을 개발함으로서 교통사고 유형 및 위험 요인 분석에 따 른 선제적 교통안전관리시스템 개발의 핵심 기술을 개발함				보고서 면수 103 면	

<국문 요약문>

연구의 목적 및 내용	[연구 목적] • 실제 수집된 다양한 교통사고 및 교통 흐름 관련 DB에 적용하여 도출된 결과를 바탕으로 제안된 기술의 성능평가 및 신뢰성 검증을 실시함 • 교통사고 이력 데이터에 교통 흐름 관련 빅데이터를 퓨전하여, 고급 통계 기법 및 데이터마이닝 (data mining) 기법을 적용함으로써, 사고요인에 대한 체계적인 분석과 이를 바탕으로 한 선제적 안전관리 기법 개발 [연구 내용] 1. 교통사고 위험요인 분석을 위한 관련 데이터 통합 기술 연구 • 국내외 활용 가능한 교통사고 관련 DB 조사 수집 및 분석 • 위험요인 분석을 위한 통합데이터 플랫폼 개발 • 데이터 통합 기술 검증 2. 교통사고 심각도 결정모델 개발 • 교통사고 심각도 분석 연구의 선행 연구 조사 • 적용 가능한 데이터마이닝 기법 조사 및 적용 사례 분석 • 교통사고 위험요인 분석을 위한 방향성 제시 • 개발된 모델의 성능평가 및 신뢰성 검증					
연구개발성과	교통 빅데이터를 기반으로 다양한 데이터 사이언스기법을 적용함으로써 교통사고 건수 등 단순한 수리적 리스크 산정이 아닌 사고결과와 사고 요인간의 연관성을 기계적 학습을 통해 사고 결과와 요인간의 유기적 관계를 정량화 함					
연구개발성과의 활용계획 (기대효과)	1. 기술적 효과 • 도로 환경 및 상황적 요인이 교통사고 발생에 미치는 영향을 분석하는 기반제공 • 자율 주행 등 미래 교통 시스템 안전관리 기반 기술 마련 2. 경제적 효과 • 빅데이터의 지속적인 성장에 따른 교통 분야의 시스템 매출 증대 • 선진국 수준의 교통사고 분석 기술 개발 3. 사회적 효과 • 보다 안전한 도로교통 시스템을 통한 국민체감적 안전확보 • 선진형 도로 교통 안전관리 문화 수립 기반 마련 • 데이터에 기반한 효율적인 사고예방과 위험경감 조치 방안 제공					
핵심어 (5개 이내)	교통사고분석	위험요인분석	빅데이터	데이터마이닝	데이터통합	

〈 SUMMARY 〉

Purpose & Contents	[Purpose] • Perform the proposed methodology's performance evaluation and reliability verification of the result based on the various collected traffic accident data and traffic flow related databases • Develop a proactive road traffic safety management methodology through systematic risk factor analysis by using advanced statistics and data mining technique on integrated big data of traffic accident history data and traffic flow data [Contents] 1. Research on data integration technique for traffic accident risk factor analysis • Survey and analysis on internal and external traffic accident related databases • Develop integrated data platform for risk factor analysis • Verify the proposed data integration methodology 2. Development of traffic accident injury severity decision model • Literature review of risk factor analysis and relation of injury severity • Investigate on applicable data mining technique and analysis of application case research • Performance evaluation and verification				
Results	Instead of applying basic statistics to calculate the mathematical risk such as the number of accidents, we applied various data science techniques based on big data in transportation to quantify the correlations between the factors and results of accidents				
Expected Contribution	1. Science & Technical Effect • Establish a foundation for analysis of impact of road environment characteristics and circumstantial to traffic accident • Contribute to the development of smart road traffic system 2. Economical Effect • Drive revenue growth of transportation system through consistent development of big data collection and its analysis • Expand to international market with traffic accident analysis technology 3. Social Effect • Procure of national degressive safety through safer road traffic system • Provide a foundation of establishing advanced road traffic culture • Provide a measure of efficient accident prevention and risk mitigation				
Keywords	Traffic accident analysis	Risk factor analysis	Big-data	Data mining	Data integration

< CONTENTS >

1. Summary of Research Projects	7
2. Domestic and International R&D Trends	16
3. Research Contents and Result	19
4. Goal Achievement and Contributions on Related fields	86
5. Future Plan of R&D Results	88
6. International Scientific and Technical Information	90
7. Security Level	91
8. Registered Research Facilities and Equipment Status on NTIS	91
9. Safeguards Implementation Performance	91
10. Representative Research Results	92
11. Etc.	93
12. References	101

〈 목 차 〉

1. 연구개발과제의 개요	7
2. 국내외 기술 개발 현황	16
3. 연구 수행 내용 및 성과	19
4. 목표 달성도 및 관련 분야 기여도	86
5. 연구개발성과의 활용 계획	88
6. 연구 과정에서 수집한 해외 과학기술 정보	90
7. 연구개발성과의 보안등급	91
8. 국가과학기술종합정보시스템에 등록한 연구시설·장비 현황	91
9. 연구개발과제 수행에 따른 연구실 등의 안전 조치 이행 실적	91
10. 연구개발과제의 대표적 연구 실적	92
11. 기타 사항	93
12. 참고 문헌	101

제1장. 연구개발과제의 개요

1. 연구개발 목적

- 본 연구는 선제적 교통안전관리시스템의 핵심 기술인 교통사고 위험요인 분석을 최종 목표로 함
- 대용량 교통사고 이력 데이터에 교통 흐름 관련 데이터를 퓨전하여 사고 데이터의 한정된 도로 환경적 정보를 보완하고, 고급 통계기법 및 데이터 마이닝 (data mining) 기법을 적용하여 교통사고의 유형 및 사고의 부상 심각도에 영향을 주는 위험요인을 파악하는 데 그 목적이 있음
- 본 연구를 통해 기존의 분석기법과 차별화된 위험요인 분석 기술을 개발하여, 실제 수집된 다양한 교통사고 DB 및 교통 흐름 관련 DB에 통합 적용함으로써 고속도로, 일반국도, 지방도 및 시도 등 관리기관에 제약 받지 않는 통합 솔루션을 개발

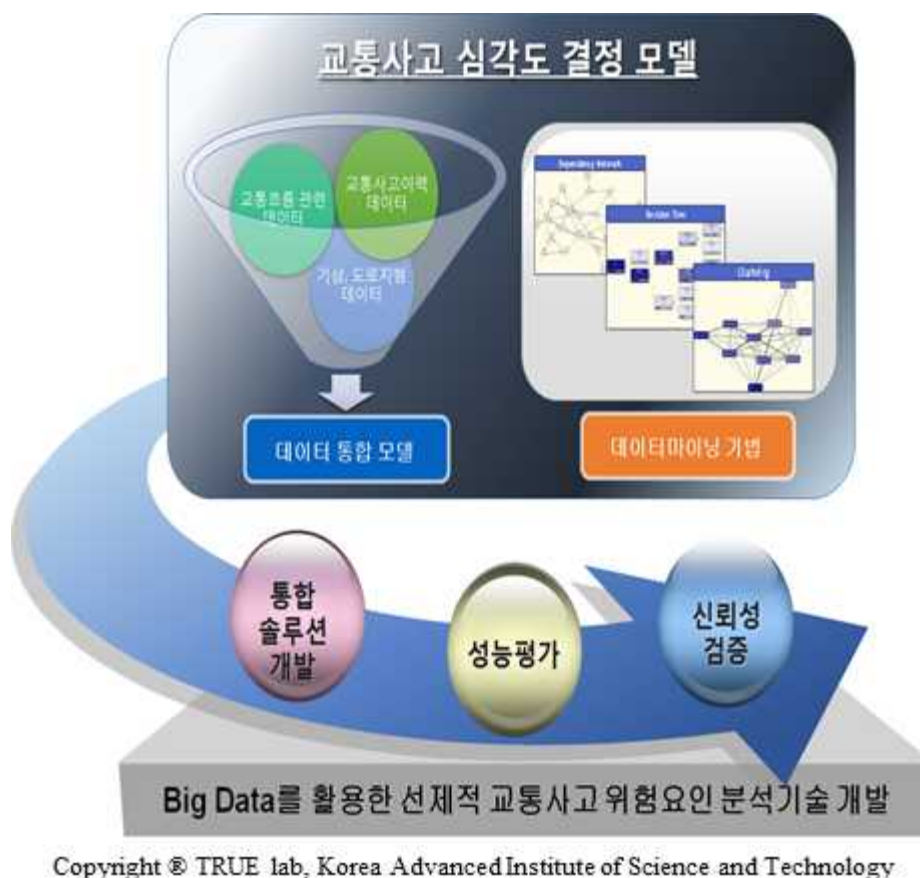


그림 1 연구 목표 구성도

2. 연구개발의 필요성

가. 연구개발과제의 설명

(1) 연구개발 개요

- 다양한 사회적 요인과 삶의 질적 향상에 맞추어 교통에 대한 수요는 꾸준한 증가세를 보이고 있으나 그와 동시에 교통사고로 인한 인적 물적 피해 또한 커다란 사회·경제적 문제로 대두되어 왔음
- 그러나 기존의 단편적인 분석을 통해서는 교통안전에 대한 효율적인 개선책을 마련하는 데 있어 많은 어려움이 드러났고, 이를 해결하기 위한 새로운 기술 개발이 시급해진 상황임
- 정부의 경우, 자료 저장 단가의 감소와 행정 디지털화를 밑바탕으로 한 기존 자료의 전산화 과정을 통하여 수많은 데이터들을 축적하기 시작했고, 이렇게 축적된 방대한 양의 데이터를 어떻게 분석하고 활용할 것인가는 교통연구 분야에서 중요한 이슈로 떠오르고 있음
- 이러한 맥락에서 본 연구는 다양한 대용량 교통사고 이력 데이터들에 고급 통계기법 및 데이터마이닝 (data mining) 기술을 적용함으로써 교통사고 유형 및 위험요인분석을 통해 선제적 교통안전관리시스템 개발의 핵심 기술을 개발하고자 함

(2) 추진배경 : 국내 교통사고 발생 현황

- 국내 교통사고 발생 현황을 살펴보면 그간 꾸준한 교통안전 확립을 위한 노력의 결과로 최근 10년간 사망자수는 감소추세를 보였으나, 부상자수에 있어서는 별다른 변화가 없었던 것으로 집계되었음
- 하지만 이를 OECD 가입 국가들을 기준으로 비교해봤을 때, 인구 10만 명당 교통사고 발생 건수와 사망자수가 각각 2011년 기준으로 445.4건, 10.5명으로 OECD회원국 중 3위, 2위를 기록함
- 그 중 교통약자인 어린이와 노인층 보행사고는 심각한 수준으로, 10만 명당 보행 사망자수가 각각 0.7명, 15.6명으로 OECD 회원국 중 4위, 1위를 기록한 바 있음. 이상의 현황들을 살펴보았을 때, 교통사고 발생 건수와 사망자 수준을 줄이기 위해 선진 교통안전관리시스템 구축이 시급한 것으로 판단됨

교통사고 발생현황(사망, 부상)



그림 2. 국내 교통사고 발생현황 (출처 : 나라지표 - www.index.go.kr)



그림 3. 인구 10만 명당 교통사고 사망자수

(출처 : <http://internationaltransportforum.org/IRTADUsers>, 13-7)

(건/10만명)



그림 4. 인구 10만 명당 교통사고 발생 건수

(출처 : <http://internationaltransportforum.org/IRTADUsers>, 13-7)

나. 연구개발의 중요성

(1) 기술 개발의 중요성

- 우리나라 교통사고 발생건수와 사망률은 OECD국가들 중 최하위 권에 머물고 있으며, 전 세계를 대상으로 확대해도 하위권에 머무르고 있음. 특히 100만 km 주행거리 당 사망자의 경우 2009년 현재 20.1명이 사망자수로 세계 1위의 불명예를 안았고, 가장 최근 통계인 2011년에서도 상위권에 자리 잡고 있어 그 대책이 시급한 실정임
- 하지만 교통안전 대책은 사후적 대응 수준에 머물고 있는 등 데이터 기반의 과학적 안전관리기술은 선진국에 비해 뒤쳐지고 있는 실정임. 따라서 사전적 대응 기법 개발 및 효율적인 개선책 마련을 위한 정확한 사고 원인과 위험요인 분석은 필수적임
- 그러나 국내의 교통안전 정책은 이러한 분석이 미비한 가운데 행정적인 절차에 의해 일률적인 개선책으로 시행되었고, 10여 년간 교통사고 사망률은 점진적으로 줄어왔지만, 발생건수는 그대로 유지되고 있는 상황이 기존의 방법론이 가지고 있는 한계를 보여주고 있음.

- 본 연구는 기존의 한계를 뛰어넘는 새로운 방법론을 통하여 효과적인 교통안전 정책 수립을 뒷받침하는 한편 효율적인 예산 집행을 가능하게 함으로써 그 중요성이 매우 크다고 할 수 있음

(2) 정부 지원의 필요성

- 기획재정부가 발표한 자료에 따르면 2012년 현재 교통안전 예산 총액은 1조 4556억 원으로 산정. 이 중 교통안전과 직접적인 상관성이 없는 비용을 제외하고 남은 금액은 약 4620억 원으로 일본의 약 4% 수준으로 지원이 열악한 상황임
- 연구 R&D예산 부문으로는 국토 교통부에서 발주하는 R&D 사업의 2013년도 예산 현황에서는 교통 물류 연구 산업이 약 566억 원이 책정되어 있으며, 이 중 일부만이 교통안전에 사용되고 있는 것으로 파악, 지원 규모의 많은 차이가 나는 것을 알 수 있음
- 또한 미국 캘리포니아 주의 경우, 2012년에서 2013년의 주 고속도로 운영에 20억 달러(약 2조 원)의 예산을 책정하였으며, 그 중 교통사고 감소를 위한 예산으로 3.5억 달러(약 3500억 원)을 할당함. 이 액수가 캘리포니아 주의 고속도로에 한한다는 점을 감안하면 우리나라 교통 안전 분야에 대한 지원이 부족한 실정이라 결론내릴 수 있음

다. 연구개발 단계(TRL에 근거하여 기술)

구분		표기	단계별 설명	
기초연구 단계	TRL 1	☐	TRL 1	이론적 연구단계
	TRL 2	☐	TRL 2	기술적 개념정립 단계
실험 단계	TRL 3	☒	TRL 3	실험적 증명 단계
	TRL 4	☐	TRL 4	실험실 성능평가 단계
시작품 단계	TRL 5	☐	TRL 5	유사환경 시험 단계
	TRL 6	☐	TRL 6	유사환경 기술시범 단계
실용화 단계	TRL 7	☐	TRL 7	실제환경 시작품 데모 단계
	TRL 8	☐	TRL 8	시제품 설치 및 성능시험 단계
사업화 단계	TRL 9	☐	TRL 9	현장적용/사업화 단계

-
- 빅데이터를 활용한 선제적 교통사고 위험요인 분석 기술은 TRL 3단계의 기술 성숙도를 지님. 빅데이터 분석을 위한 데이터 마이닝 기술에 관한 개념적 정립이 확립되어 가는 동향에서, 본 기술은 이러한 기술을 적용한 퓨전된 교통 안전 데이터베이스의 해석적 연구를 통하여 분석 기술을 확보하는데 주력함

라. 타산업 연계 및 국제협력 가능성

(1) 타 산업 연계 가능성

- 본 연구에 사용되는 각종 빅데이터 및 데이터 마이닝 관련 기술의 최대 장점은 지능형 개발산업에서 매우 큰 비중을 차지하는 기술임. 대표적 분야로 지능형 교통시스템(ITS)이 있음.
- 내비게이션 어플리케이션 개발 등 교통정보를 기반으로 한 지식서비스 제공 산업부터 지능형 교통정보 수집 센서 제작 산업, 신호 제어 및 통신 등 다양한 업종들의 경계를 넘나들 수 있음.
- 국토교통부의 교통정보 플랫폼 포럼 창립 등 ITS 산업 활성화의 흐름 속에서 공기업과 민간기업의 구분 없이 연계, 활용될 것임. 뿐만 아니라, 지능형 교통시스템을 효율적으로 이용할 수 있는 물류 업계 및 운수 업계와도 연계 가능함

(2) 국제 협력 가능성

- 본 연구에서는 미국을 포함한 선진국들의 우수 연구 기관들과 연구 협력을 추진해 나갈 계획임. 이들 연구 기관들은交通安全 관련 연구에 있어서 다양한 기법들을 연구 개발해왔으며 전 세계적으로 이 분야를 선도하고 있어 이 분야에 대한 노하우를 공유할 수 있을 것으로 기대됨
- 현재 본 연구진은 미국의 우수 연구 기관과 연구 협약 MOU를 체결했으며, 이외에도 캘리포니아 교통청 등 여러 정부 기관들과 다각적인 연구 협력 관계를 통해 다양한 공동연구를 진행 중임

3. 연구개발 범위

가. 연구개발 최종목표

(1) 최종목표의 성격

- 본 연구는 교통안전증진의 핵심기술 개발을 목표로 빅데이터를 활용한 선제적 교통사고 위험요인 분석을 연구 최종 성과물로 함
- 최종 연구개발의 목표 달성을 위해 세부 성과물로는 활용 가능한 교통 DB 조사 목록과 서로 다른 구조의 DB를 퓨전하기 위한 DB 통합 모델, 그리고 교통사고 심각도 결정 모델 개발이 있음

(2) 설정근거.

- 획기적인 교통안전증진을 위해선 선제적 교통사고예방과 체계적인 교통안전관리시스템 개발이 필수적이거나 그에 앞서 체계적인 교통안전관리를 하는 데 있어 핵심적 역할을 수행하는 위험관리 부분의 분석이 시급한 실정임
- 그러나 교통사고 분석에서 위험요인들 간의 복잡한 상관관계와 잠재적인 위험요인들로 인하여 사고의 심각도에 영향을 주는 주요 위험요인 파악 및 정확한 교통사고 심각도 결정 모델 개발은 단순 통계분석으로는 불가능하며, 고급 통계기법 및 비매개 모델 등과 같은 기술을 요구함
- 현 시점에서의 데이터 마이닝 기법을 활용한 위해 요인 분석 기술은 유사 연구조차 제대로 이루어지지 않은 국내 상황으로 인하여 응용 단계의 기술적 성숙도에 따른 선행연구를 요구함. 이러한 선행 연구에는 아직까지 체계적으로 구축되지 않은 여러 관련 데이터베이스를 조사하고 분석 가능성을 타진하는 것이 필수적일 것으로 판단됨
- 이러한 선행연구에 대한 수요와 한정된 연구 자원을 고려해볼 때, 본 연구에서는 교통사고 심각도를 결정하는 위험요인을 파악하고 평가하는 교통사고 위험요인 분석을 최종 목표로 설정함

나. 연차별 연구목표, 내용 및 범위

(1) 연구목표 및 내용

연구 내용			
A-1	교통사고 위험요인 분석을 위한 관련 데이터 통합 기술	A-1-①	활용 가능 데이터베이스 조사
		A-1-②	통합 데이터 플랫폼 개발
		A-1-③	데이터 통합기술 검증
A-2	교통사고 심각도 결정 모델	A-2-①	사고 심각도 분석 선행 연구 조사
		A-2-②	위험요인 분석을 위한 방향성 제시
		A-2-③	데이터 마이닝 기법을 활용한 사고 심각도 결정 모델 개발
		A-2-④	정확도 산출 수치를 활용한 기존 모델과의 성능 비교

(2) 연구개발 범위

연구범위	연구수행방법 (이론적·실험적 접근 방법)
국내외 활용 가능한 교통사고 관련 DB 조사 수집 및 분석	- 국내외 경찰 및 보험사 등의 교통사고 DB 사전 조사 분석 - 국내외 차량 검지기 (VDS), 고속도로 통행료 수납 시스템 (TCS), 단거리전용 통신 장비 (DSRC) 등의 교통량, 속도, O/D관련 DB 및 그 밖의 기상, 교통지형 DB 사전 조사 분석 - 특성별 교통사고 DB 조사 목록 작성 - 위험요인분석기술 개발을 위한 분석대상자료 선정 - 분석대상 자료 확보를 위해 우선적으로 최근 정부의 공공 데이터 개방정책에 따라 정부기관이나 지방자치단체가 제공하는 다양한 교통 관련 데이터를 활용 - 원자료(raw data)나 상세 항목을 포함한 분석대상 자료의 수집 및 확보를 위해 관련 기관 및 정부 부처와의 연구협

	력관계 체결 및 공문발송 등의 정식 절차를 수행 - 분석대상 자료의 보완 및 자료에 대한 정확한 이해를 위해 관련 기관들과의 긴밀한 협조를 유지하고, 정보공유 차원에서 지속적인 교류와 세미나 등을 통한 공조를 이루어 나감 - MOU체결이 맺어진 해외 연구기관들과의 연구협력을 통해 국외의 교통 관련 DB 분석 및 확보
대용량 데이터 통합 기법 및 관련 연구 조사	- 국내외 문헌 조사를 통한 데이터 퓨전, 통합, 취합 기술의 활용 사례 조사 - 데이터 통합 모델 개발에 관한 기술동향 및 구축사례 사전 조사
교통사고 심각도 분석 연구의 선행 연구 조사	- 로지트 (logit) 및 프로비트 (probit) 회귀분석을 활용한 기존 교통사고 심각도 분석에 관한 선행 연구 조사 - 교통사고 분석 연구의 기술 동향 분석 - 해외 우수 연구기관과의 연구 협력을 통한 선진 연구 업적 및 노하우 습득
적용 가능한 데이터마이닝 기법 조사 및 적용 사례 분석	- 연관성 분석 (association), 분류 분석 (classification), 군집 분석 (clustering) 등에 해당하는 데이터마이닝 알고리즘 조사 및 이론적 연구 - 범주형 데이터 분석을 위해 적용 가능한 비매개 분석 기법 조사 및 사례 분석
교통사고 위험요인 분석을 위한 방향성 제시	- 선행 연구 조사 및 사례 조사 분석을 통한 연구 방향성 설정 및 타당성 검토

제2장. 국내외 기술 개발 현황

1. 관련 분야의 국내외 연구동향 및 학술적 성과

가. 국가 연구개발사업 유사연구 실적

(1) 연구개발과제의 중복방지를 위한 조사 및 검토결과

- 국가과학기술지식정보서비스의 유사과제 중복성 검토 결과 기수행과제 및 타인등록과제 상에서 유사과제는 검색되지 않았으나, 세부과제 검색 결과 중복 가능성이 있는 과제는 다음과 같음

(단위 : 억원)

부처명	사업명	과제명	예산	비 고
교육과학기술부	일반연구자 지원사업	공간계량분석을 통한 도시교통사고예측 및 사고심각도 분석 모형개발	0.56	2012.09~2015.08

(가) 유사 과제의 연구 내용

- GIS기반 사고 데이터 수집 및 교통시설, 토지이용형태, 사회경제지표, 교통안전정책변수와 교통사고와의 관계분석
- 교통사고발생에 대한 공간적 자기상관 검증과 공간이 가지는 가변적인 요인에 대한 계량적 추정 수행을 통한 공간계량분석적 도시교통 사고예측 모형 및 도시교통 사고심각도 분석 모형 개발
- 지역단위의 교통안전 정책 수립 및 평가를 위한 가이드라인 제시

(2) 연구개발과제의 차별화 방안

- 본 과제에서는 기존 과제의 지역단위별 거시적 모형과 차별화된 미시적 개념의 개별 사고 기반의 위험요인 분석을 수행함. 따라서 도시교통사고로 국한된 형태의 모형이 아닌 고속도로, 국도, 시도 등 모든 교통사고에 적용 및 분석 가능한 통합 솔루션 개발이 이루어짐

-
- 또한 특정한 함수형태와 매개변수를 요구하는 계량모형과 달리 방대한 양의 데이터에서 변수들 간의 상관관계를 고려하여 신속히 결과를 도출 해 낼 수 있는 비매개 분석 기법을 사용하여 좀 더 정확하고 우수한 성능의 사고 심각도 결정 모델을 개발하는 것이 목적으로, 유사 과제와 분명한 차이를 보임

2. 국내외 시장분석 및 산업동향

가. 시장분석

- 시장 조사기관 IDC는 세계 빅데이터 시장이 매년 39~60% 성장하여 2015년에는 169~321억 달러 규모로 증가할 것으로 전망함 교통 분야는 사회 곳곳에 활용되고 있는 빅데이터 분석에서 가장 우선시 되는 적용 영역 중에 하나로, 최근 일본은 실시간 교통상황 데이터를 종합 분석해 지능형 교통안내 시스템을 개발하여 교통 체증 감소 효과를 누리고 있는 상황임.
- 또한 독자적 빅데이터 분석처리기술을 보유한 IBM, ORACLE, GOOGLE, MS 등 글로벌 IT기업들은 이미 ‘12년부터 빅데이터 시장 선점을 위해 관련 서비스를 출시하고 있으며 교통 흐름 분석 등에 사용 중임. 국내의 SK C&C, LG CNS, 삼성SDS도 ‘13년부터 빅데이터 분석 전문 인력을 2~5배까지 늘여 독자적 솔루션 플랫폼을 개발하고 있고, SK C&C는 현재 실시간 대용량 데이터 분석 솔루션으로 교통 분석을 제공하고 있음

나. 산업 동향

- 도로교통 공단은 새 정부의 핵심 국정과제인 정부 3.0의 확산을 돕는 차원에서 UTIS 빅데이터를 활용한 도로교통공단의 도로교통 3.0 과제를 선정하여 추진하고 있음. 뿐만 아니라 청주시의 교통정보 빅데이터 분석시스템이 수립되는 등 여러 지자체들의 정책적인 지원이 뒷받침 되고 있으며, 차후 응용이 가능한 스마트 인프라 시스템 구축과 관련하여 LG, IBM 여러 유수 기업 등이 제주도 등 여러 지자체와 연계 사업을 진행하고 있기에 앞으로의 연구 및 시장성이 주목됨
- 또한 국토교통부는 ‘16년까지 빅데이터와 공간정보를 융합한 공간 빅데이터를 구축하여 교통 혼잡도 분석 등에 사용할 예정이며, 서울시와 서울지방경찰청은 2020년까지 빅데이터를 활용하여 교통사고 사망자 수 절반으로 줄이기 위한 개선사업을 진행 중임

-
- 서울시에서는 ‘13년부터 데이터 분석 기반을 마련하기 위해 정책 과제를 추진했으며, 올해 개발된 기술의 우선 시범 적용 분야로 교통을 선정, 다양한 교통현안 해결에 활용 가능할 것이라 예상함

제3장. 연구 수행 내용 및 성과

1. 교통사고 위험요인 분석을 위한 관련 데이터 통합 기술 연구

가. 국내외 활용 가능한 교통사고 관련 데이터베이스 조사 수집 및 분석

- 1) 국내외 교통사고 관련 데이터베이스는 다양한 테이블 및 데이터 항목으로 구성되어 있으며 데이터가 수집되는 경로 및 방법 또한 다양함
- 2) 특정 데이터베이스만을 이용한 교통사고 위험요인 분석도 가능하지만 다양한 데이터베이스를 통합하여 분석 시 다양한 교통사고 위험요인을 밝혀낼 수 있을 것으로 기대
- 3) 데이터 통합을 위해서는 개별 데이터베이스의 테이블, 테이블 별 데이터 항목, 데이터 수집 경로 및 방법에 등에 대한 이해가 필요
- 4) 다양한 교통사고 관련 데이터베이스에 대한 이해를 위해 SWITRS, PeMS, TAAS 등 국내외의 다양한 교통사고 관련 데이터베이스를 수집하고 분석함

가) Caltrans Traffic Safety Investigation

- (1) 캘리포니아 교통국 Caltrans에서는 캘리포니아 고속도로로부터 수집된 데이터를 이용하여 교통을 분석하는 Traffic Investigation Report Tracking System (TIRTS)를 운영하고 있음
- (2) 교통조사는 Traffic Safety Investigation, Traffic Operational Investigation, 그리고 Americans with Disabilities Act (ADA) Accessibility Investigation의 세 가지 하위 분야로 나뉘며, 이 중 사고 다발지역 선정은 Traffic Safety Investigation (교통안전 전 조사)에서 이루어짐
- (3) Traffic Safety Investigation은 Table C (All), Table C (Wet), Other Safety의 세 가지 조사를 바탕으로 이루어짐
- (4) 이 중 Table C 조사의 경우 자동화된 시스템으로 주 평균 사고율보다 현저하게 높은 사고율을 보이는 고속도로 구간을 찾아내어 Table C List에 기록
- (5) Other Safety Investigation의 경우 교통국에서 실시하는 연례 실사 점검, 혹은 민원에 의해 이루어짐
- (6) Caltrans의 교통조사보고서 (TIR)에서는 분기별로, 과거 3년 치에 해당하는 사고 데이터를 이용한 Table C 조사를 통해 사고 다발지역 (HCCL, High Collision Concentration Location) 선정이 이루어짐

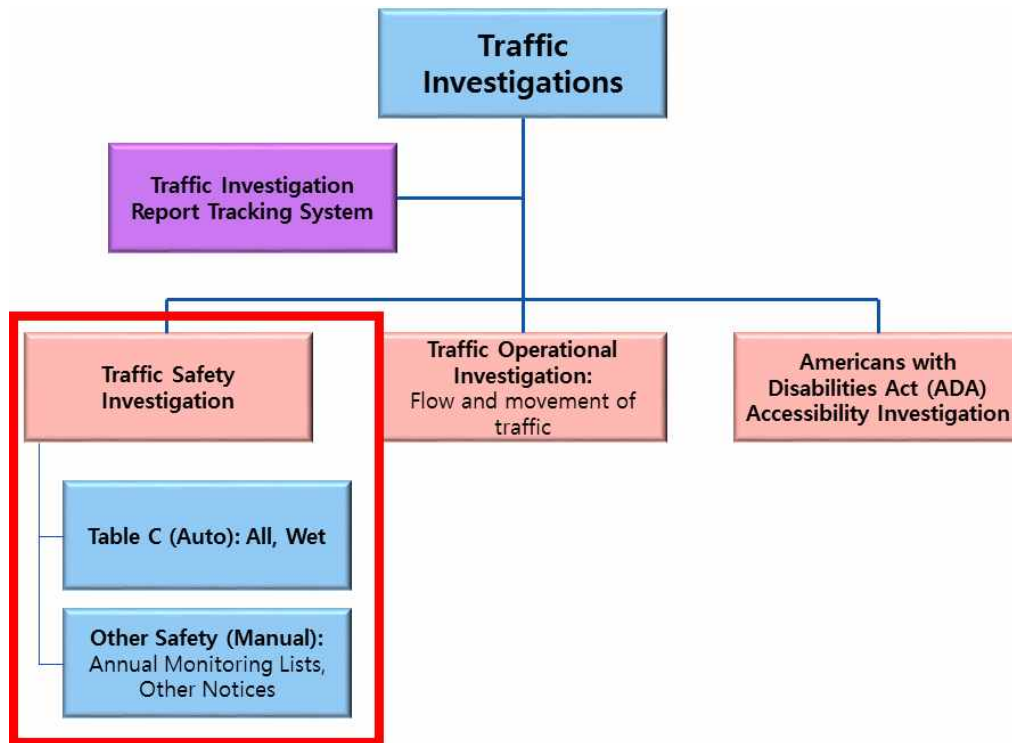


그림 6 Caltrans Traffic Safety Investigation의 구조

(7) Table C를 이용한 사고다발지역 선정

(가) Table C 조사 결과를 이용하여, 다음과 같이 사고 다발 지역에 대한 정보를 얻을 수 있음

- ① 대상 도로 정보(구역, 고속도로, 번호, 방향), 구간 시작 위치, 구간 종료 위치, 길이, 구간 내 사고 건수
- ② 다음 표는 2008년에 기록된 Table C 조사 결과를 바탕으로 선정한 4구역 (District 4)의 서쪽 방향 I80 고속도로상의 사고다발지역의 일부를 나타낸 것임

No. of HCCL	Table C	Starting Postmile	Ending Postmile	Length	Accident count
1	2008/1 st	0.98	1.18	0.20	23
2	2008/1 st	1.50	1.70	0.20	28
3	2008/1 st	1.90	2.10	0.20	45
4	2008/2 nd	2.10	2.30	0.20	16
5	2008/3 rd	2.94	3.14	0.20	15
6	2008/1 st	4.08	4.48	0.40	58
7	2008/3 rd	5.51	5.71	0.20	9
...	...	...	...	...	...

그림 7 사고다발지역 예시

나) SWITRS (Statewide Integrated Traffic Records System)

- (1) 미국 캘리포니아에서 발생하는 모든 교통사고를 캘리포니아 고속도로 순찰대 (CHP, California Highway Patrol)가 기록해둔 대용량 데이터베이스로 1973년부터 기록됨
- (2) Crash, Party, Victim의 3개의 Table로 구성되어 있으며 계층적 구조를 이루고 있음
 - Crash- 사고에 대한 전반적 정보 (사고 위치, 사고 유형, 도로 및 기상 환경 등)를 담고 있으며 74개 데이터 항목으로 구성되어 있음
 - Party - 사고와 연루된 차량 (혹은 보행자) 및 운전자에 대한 정보로 33개 데이터 항목으로 구성되어 있음
 - Victim - 사고와 연루된 피해자들에 대한 정보로 10개의 데이터 항목으로 구성되어 있음

Crash Table

caseid	year	procdat	juris	date	time	badge	juridist	dayweek	shift	pop	location	...
4392008	2010	2010-02-19	1942	2010-01-03	1405	38876	527	7	S	7	1942	...
4392009	2010	2010-02-24	9855	2010-02-06	450	19462		6	3	6	3616	...
4392010	2010	2010-02-19	1005	2010-01-24	1909	P904	CE245	7	S	7	1005	...
4392011	2010	2010-02-19	1005	2010-01-15	135	P731	NW144	5	S	7	1005	...
4392012	2010	2010-02-19	1900	2010-01-31	2123	462031	560	7	S	3	1968	...
4392013	2010	2010-02-09	9835	2010-01-17	1835	16200		7	2	3	3601	...
...	...	...	...	...	...	...	...	...	...	...	...	...

Party Table

caseid	parnum	pptype	atfault	psex	page	psober	pdrug	pdirect	psafety1	psafety2	insured	...
4392009	1	1	Y	M	51	A	-	S	L	G	Y	...
4392009	2	1	N	M	25	A	-	S	M	G	Y	...
4392009	3	1	N	F	29	A	-	S	M	G	Y	...
4392010	1	1	Y	M	45	A	-	W	M	G	Y	...
4392010	2	2	N	M	79	A	-	N	-	-	O	...
4392011	1	1	Y	M	51	B	-	N	P	W	Y	...
...	...	...	...	...	...	...	...	...	...	...	...	...

Victim Table

caseid	parnum	vtype	vsex	vage	vinjury	vseat	vsafety1	vsafety2	vejected
4392009	1	1	M	51	3	1	L	G	0
4392009	2	1	M	25	1	1	M	G	0
4392009	3	2	F	29	0	1	M	G	0
4392010	1	2	F	38	0	3	L	G	0
4392010	2	3	M	79	1	9	-	-	3
...	...	...	...	...	...	...	...	...	...

그림 8 SWITRS 데이터베이스 구조

(3) Crash 테이블

(가) Crash 테이블의 주요 데이터항목은 아래와 같음 (04' -10')

① 인간의 행동에 따라 분류된 데이터 항목과 사고의 분포

Data Field	Values	Collision Count	Proportion, %
PCF: Primary Collision Factor	Not Stated	24638	0.73%
	A - (Vehicle) Code Violation	3472824	93.66%
	B - Other Improper Driving	22002	0.65%
	C - Other Than Driver	69078	2.04%

	D - Unknown	97116	2.87%
	E -Feel Asleep	1785	0.05%
VIOLCAT: Primary Violation Category	00 - Unknown	103801	3.06%
	01 - Driving or Bicycling Under the Influence of Alcohol or Drug	251083	7.41%
	02 - Impeding Traffic	1925	0.06%
	03 - Unsafe Speed	1026970	30.32%
	04 - Following Too Closely	76950	2.27%
	05 - Wrong Side of Road	84864	2.51%
	06 - Improper Passing	31814	0.94%
	07 - Unsafe Lane Change	225451	6.66%
	08 - Improper Turning	567824	16.76%
	09 - Automobile Right of Way	413872	12.22%
	10 - Pedestrian Right of Way	31327	0.92%
	11 -Pedestrian Violation	34181	1.01%
	12 - Traffic Signals and Signs	185507	5.48%
	13 - Hazardous Parking	4456	0.13%
	14 - Lights	498	0.01%
	15 - Brakes	1026	0.03%
	16 - Other Equipment	4648	0.14%
	17 - Other Hazardous Violation	36973	1.09%
	18 - Other Than Driver (or Pedestrian)	69078	2.04%
	21 - Unsafe Starting or Backing	135159	3.99%
	22 - Other Improper Driving	22002	0.65%
	23 - Pedestrian or “Other” Under the Influence of Alcohol or Drug	2	0.00%
	24 - Fell Asleep	1785	0.05%
HITRUN: Hit And Run	F - Felony	90184	2.66%
	M - Misdemeanor	564035	16.65%
	N - Not Hit and Run	2733224	80.69%
PED: Pedestrian	Not Stated	1333	0.04%
	A - No Pedestrian Involved	3289710	97.11%

Action	B - Crossing in Crosswalk at Intersection	40693	1.20%
	C - Crossing in Crosswalk Not at Intersection	2172	0.06%
	D - Crossing Not in Crosswalk	28282	0.83%
	E - In Road, Including Shoulder	17677	0.52%
	F - Not in Road	7467	0.22%
	G - Approaching/Leaving School Bus	109	0.00%
ETOH:	N - No	3041107	89.78%
Alcohol Involved	Y - Yes	346336	10.22%
Overall		3387443	100%

② 차량 특성에 따라 분류된 데이터 항목과 사고의 분포

Data Field	Values	Collision Count	Proportion, %
STFAULT: Statewide Vehicle Type At Fault	Not Stated	865151	25.54%
	A - Passenger Car/Station Wagon	1852943	54.70%
	B - Passenger Car with Trailer	6461	0.19%
	C - Motorcycle/Scooter	50698	1.50%
	D - Pickup or Panel Truck	381275	11.26%
	E - Pickup or Panel Truck with Trailer	15485	0.46%
	F - Truck or Truck Tractor	38447	1.13%
	G - Truck or Truck Tractor with Trailer	51473	1.52%
	H - Schoolbus	6199	0.18%
	I - Other bus	9854	0.29%
	J - Emergency Vehicle	15495	0.46%
	K - Highway Construction Equipment	727	0.02%
	L - Bicycle	49636	1.47%
	M -Other Vehicle	10307	0.30%
	N - Pedestrian	32967	0.97%
	O - Moped	325	0.01%
Overall		3387443	100%

③ 도로 특성에 따라 분류된 데이터 항목과 사고의 분포

Data Field	Values	Collision Count	Proportion, %
ROADSURF: Road Surface	Not Stated	36497	1.08%
	A - Dry	3022538	89.23%
	B - Wet	307717	9.08%
	C - Snowy or Icy	16760	0.49%
	D - Slippery (Muddy, Oily, etc.)	3931	0.12%
RDCOND1: Primary Road Condition	Not Stated	34999	1.03%
	A - Holes, Deep Ruts	11954	0.35%
	B - Loose Material on Roadway	7554	0.22%
	C - Obstruction on Roadway	14023	0.41%
	D - Construction or Repair Zone	48590	1.43%
	E - Reduced Roadway Width	3750	0.11%
	F - Flooded	3070	0.09%
	G - Other	14095	0.42%
	H - No Unusual Condition	3249408	95.93%
RDCOND2: Secondary Road Condition	Not Stated	3374467	99.62%
	B - Loose Material on Roadway	249	0.01%
	C - Obstruction on Roadway	154	0.00%
	D - Construction or Repair Zone	1306	0.04%
	E - Reduced Roadway Width	2776	0.08%
	F - Flooded	137	0.00%
	G - Other	604	0.02%
	H - No Unusual Condition	7750	0.23%
Overall		3387443	100%

④ 연도 및 요일에 따라 분류된 데이터 항목 및 사고의 분포

Data Field	Values	Collision Count	Proportion, %
YEAR: Collision Year	2004	538295	15.89%
	2005	532725	15.73%
	2006	519203	15.33%
	2007	501907	14.82%
	2008	452595	13.36%
	2009	426228	12.58%
	2010	416490	12.30%
DAYWEEK: Day of Week	1 - Monday	471900	13.93%
	2 - Tuesday	495463	14.63%
	3 - Wednesday	494164	14.56%
	4 - Thursday	496023	14.64%
	5 - Friday	561836	16.59%
	6 - Saturday	469598	13.86%
	7 - Sunday	398459	11.76%
Overall		3387443	100%

⑤ 사고 발생 지점 주변 환경에 따라 분류된 데이터 항목 및 사고의 분포

Data Field	Values	Collision Count	Proportion, %
POP: Population	0 - University (Private Property)	1659	0.05%
	1 - Incorporated (less than 2500)	16364	0.48%
	2 - Incorporated (2500 - 10000)	49178	1.45%
	3 - Incorporated (10000 - 25000)	144361	4.26%
	4 - Incorporated (25000 - 50000)	304035	8.98%
	5 - Incorporated (50000 - 100000)	624971	18.45%
	6 - Incorporated (100000 - 250000)	638458	18.85%
	7 - Incorporated (over 25000)	836136	24.68%
	9 - Unincorporated (Rural)	772281	22.80%
INTERSECT: Intersection	Not Stated	46032	1.36%
	N - Not Intersection	2553150	75.37%
	Y - State Highway	788261	23.27%
STATEHW:	Not Stated	532	0.02%

State Highway Indicator	N – Not State Highway	2150400	63.48%
	Y – State Highway	126511	36.50%
CALTRANS: California County	Not Stated	2155859	63.64%
	LA – Los Angeles County	312565	9.22%
	ORA – Orange County	112579	3.32%
	SBD – San Bernardino County	82079	2.42%
	SD – San Diego County	80816	2.38%
	RIV – Riverside County	69886	2.06%
	ALA – Alameda County	66667	1.96%
	Other Counties	506992	14.97%
STROUTE: State Route	Not Stated	2155859	63.64%
	101 – US 101 Route	116920	3.45%
	5 – I-5 Route	114909	3.39%
	10 – I-10 Route	70835	2.09%
	405 – I-405 Route	53039	1.57%
	99 – I-99 Route	50906	1.50%
	80 – I-80 Route	49618	1.46%
	Other Routes	775357	22.89%
RAMP: Ramp Intersection	Not Stated	3143067	92.79%
	1 – Ramp Exit, Last 50 Feet	52215	1.54%
	2 – Mid-Ramp	56056	1.65%
	3 – Ramp Entry, First 50 Feet	10201	0.30%
	4 – Not State Highway, Ramp-related, Within 100 Feet	54994	1.62%
	5– Intersection	51851	1.53%
	6 – Not State Highway, Intersection-related, Within 250 Feet	19059	0.56%
Overall		3387443	100%

⑥ 기타 요인들에 따라 분류된 데이터 항목 및 사고의 분포

Data Field	Values	Collision Count	Proportion, %
WEATHER1: Primary Weather	Not Stated	20647	0.61%
	A - Clear	2728237	80.54%
	B - Cloudy	482808	14.25%
	C - Raining	129930	3.84%
	D - Snowing	5174	0.15%
	E - Fog	15681	0.46%
	F - Other	3667	0.11%
	G - Wind	1299	0.04%
WEAHTER2: Secondary Weather	Not Stated	3283966	96.95%
	B - Cloudy	6611	0.20%
	C - Raining	79653	2.35%
	D - Snowing	2636	0.08%
	E - Fog	2570	0.08%
	F - Other	2577	0.08%
	G - Wind	9430	0.28%
LIGHTING: Lighting	Not Stated	24626	0.73%
	A - Daylight	2276527	67.20%
	B - Dusk - Dawn	114082	3.37%
	C - Dark - Street Lights	700413	20.68%
	D - Dark - No Street Lights	265930	7.85%
	E - Dark -Street Lights Not Functioning	5865	0.17%
RIGHTWAY: Control Device	Not Stated	20355	0.60%
	A - Functioning	1147997	33.89%
	B - Not Functioning	9023	0.27%
	C - Obscured	1766	0.05%
	D - None	2208302	65.19%
Overall		3387443	100%

⑦ 사고 유형에 따라 분류된 데이터 항목

Data Field	Values	Collision Count	Proportion, %
PARTIES: Involved Party Count	1	648656	19.15%
	2	2306853	68.10%
	3	341101	10.07%
	4	70554	2.08%
	5	15050	0.44%
	Over 6	5229	0.15%
CRASHTYP: Type of Collision	Not Stated	28359	0.84%
	A - Head-On	148680	4.39%
	B - Sideswipe	283073	17.21%
	C - Rear End	1109153	32.74%
	D - Broadside	694850	20.51%
	E - Hit Object	553069	16.33%
	F - Overturned	99749	2.94%
	G - Vehicle/Pedestrian	86096	2.54%
	H - Other	84414	2.49%
INVOLVE: Involved Object With Motor Vehicle	Not Stated	18202	0.54%
	A - Non-Collision	100300	2.96%
	B - Pedestrian	93295	2.75%
	C - Other Motor Vehicle	2207495	65.17%
	D - Motor Vehicle on Other Roadway	21326	0.63%
	E - Parked Motor Vehicle	250052	7.38%
	F - Train	1096	0.03%
	G - Bicycle	77469	2.29%
	H - Animal	13123	0.39%
	I - Fixed Object	535676	15.81%
	J - Other Object	69409	2.05%
PEDCOL: Pedestrian Collision	N - No	3289720	97.12%
	Y - Yes	97723	2.88%

BICCOL:	N - No	3299435	97.40%
Bicycle Collision	Y - Yes	88008	2.60%
MCCOL:	N - No	3299235	97.40%
Motorcycle Collision	Y - Yes	88208	2.60%
TRUCKCOL:	N - No	3198349	94.42%
Truck Collision	Y - Yes	189094	5.58%
Overall		3387443	100%

⑧ 사고 심각도에 따라 분류된 데이터 항목

Data Field	Values	Collision Count	Proportion, %
TOWAWAY: Vehicle Tow Away	Not Stated	17646	0.52%
	N - No	1663740	49.11%
	Y - Yes	1706057	50.36%
CRASHSEV: Injury Severity	0 - PDO (Property Damage Only)	2091973	61.76%
	1 - Fatal	23311	0.69%
	2 - Injury (Severe)	72561	2.14%
	3 - Injury (Other Visible)	406318	11.99%
	4 - Injury (Complaint of Pain)	793280	23.42%
KILLED: Killed Victim Count	0	3364132	99.31%
	1	21423	0.63%
	2	1493	0.04%
	Over 3	395	0.01%
INJURED: Injured Victim Count	0	2105028	62.14%
	1	909472	26.85%
	2	253755	7.49%
	3	77362	2.28%
	Over 4	41826	1.23%
Overall		3387443	100%

(4) Party 테이블

(가) Party 데이터의 주요 데이터항목은 아래와 같음 (04' -10')

① 인간의 행동에 따라 분류된 데이터 항목 및 분포

Data Field	Values	Party Count	Proportion, %
ATFAULT: At Fault	N - No	3617390	54.17%
	Y - Yes	3060136	45.83%
PSOBER: Party Sobriety	Not Stated	173547	2.60%
	A - Had Not Been Drinking	5118486	76.65%
	B - Had Been Drinking, Under Influence	260059	3.89%
	C - Had Been Drinking, Not Under Influence	42986	0.64%
	D - Had Been Drinking, Impairment Unknown	52318	0.78%
	G - Impairment Unknown	609136	9.12%
	H- Not Applicable	420994	6.30%
PDRUG: Party Drug Physical	Not Stated	5580507	83.57%
	E - Under Drug Influence	25139	0.38%
	F - Impairment - Physical	9823	0.15%
	G - Impairment - Unknown	609683	9.13%
	H - Not Applicable	422042	6.32%
	I - Sleepy/Fatigued	30332	0.45%
PSAFETY1: Primary Party Safety Equipment	Not Stated	1250781	18.73%
	A - None in Vehicle	48834	0.73%
	B - Unknown	149080	2.23%
	C - Lap Belt Used	13050	0.20%
	D - Lap Belt Not Used	1836	0.03%
	E - Shoulder Harness Used	9403	0.14%
	F - Shoulder Harness Not Used	1419	0.02%
	G - Lap/Shoulder Harness Used	830986	12.44%
	H - Lap/Shoulder Harness Not Used	7761	0.12%

	J – Passive Restraint Used	997	0.01%
	K – Passive Restraint Not Used	514	0.01%
	L – Air Bag Deployed	625596	9.37%
	M – Air Bag Not Deployed	3163765	47.38%
	N – Other	92699	1.39%
	P – Not Required	448856	6.72%
	Q – Child Restraint in Vehicle Used	288	0.00%
	R – Child Restraint in Vehicle Not Used	88	0.00%
	S – Child Restraint in Vehicle, Use Unknown	339	0.01%
	T – Child Restraint in Vehicle, Improper Use	202	0.00%
	U – No Child Restraint in Vehicle	2181	0.03%
	V – Driver, Motorcycle Helmet Not Used	1753	0.03%
	W – Driver, Motorcycle Helmet Used	11956	0.18%
	X – Passenger, Motorcycle Helmet Not Used	120	0.00%
	Y – Passenger, Motorcycle Helmet Used	15022	0.22%
PSAFETY2: Secondary Party Safety Equipment	Not Stated	2191034	32.81%
	A – None in Vehicle	5790	0.09%
	B – Unknown	147928	2.22%
	C – Lap Belt Used	37979	0.57%
	D – Lap Belt Not Used	3816	0.06%
	E – Shoulder Harness Used	16831	0.25%
	F – Shoulder Harness Not Used	4571	0.07%
	G – Lap/Shoulder Harness Used	4090370	61.26%
	H – Lap/Shoulder Harness Not Used	35264	0.53%

	J – Passive Restraint Used	1539	0.02%
	K – Passive Restraint Not Used	375	0.01%
	L – Air Bag Deployed	19209	0.29%
	M – Air Bag Not Deployed	36004	0.54%
	N – Other	1545	0.02%
	P – Not Required	6226	0.09%
	Q – Child Restraint in Vehicle Used	797	0.01%
	R – Child Restraint in Vehicle Not Used	136	0.00%
	S – Child Restraint in Vehicle, Use Unknown	488	0.01%
	T – Child Restraint in Vehicle, Improper Use	328	0.00%
	U – No Child Restraint in Vehicle	305	0.00%
	V – Driver, Motorcycle Helmet Not Used	4540	0.07%
	W – Driver, Motorcycle Helmet Used	70298	1.05%
	X – Passenger, Motorcycle Helmet Not Used	49	0.00%
	Y – Passenger, Motorcycle Helmet Used	2104	0.03%
CELL: Cellphone Use	Not Stated	1584015	23.72%
	1 – Cellphone Handheld in Use	36953	0.55%
	2 – Cellphone Hands Free in Use	69032	1.03%
	3 – Cellphone Not in Use	4711322	70.55%
	B – Cellphone in Use	11720	0.18%
	D – No Cellphone/Unknown	264484	3.96%
PVIOLCAT: Violation Category in Other Associated Factor	Not Stated	6201429	92.87%
	01 – Under Influence in Public	31	0.00%
	19 – Hit and Run	22456	0.34%
	20 – Driving or Bicycling Under the	12850	0.19%

	Influence of Alcohol or Drug		
	21 - Improper Lane Change	470	0.01%
	22 - Impeding Traffic	3101	0.05%
	23 - Failure to Heed Stop Signal	9527	0.14%
	24 - Failure to Heed Stop Sign	7399	0.11%
	25 - Unsafe Speed	160518	2.40%
	26 - Reckless Driving	1251	0.02%
	27 - Wrong Side of Road	28762	0.43%
	28 - Unsafe Lane Change	16914	0.25%
	29 - Improper Passing	4155	0.06%
	30 - Following Too Closely	13026	0.20%
	31 - Improper Turning	126928	1.90%
	33 - Automobile Right-of-Way	17052	0.26%
	34 - Pedestrian Right-of-Way	1495	0.02%
	35 - Pedestrian Violation	3818	0.06%
	38 - Hazardous Parking	2386	0.04%
	39 - Lights	2920	0.04%
	40 - Brakes	1081	0.02%
	43 - Other Equipment	3696	0.06%
	44 - Other Hazardous Movement	9723	0.15%
	46 - Improper Registration	36	0.00%
	47 - Other Non-Moving Violation	16960	0.25%
	48 - Excessive Smoke	2	0.00%
	49 - Excessive Noise	4	0.00%
	50 - Overweight	16	0.00%
	51 - Oversize	194	0.00%
	53 - Unsafe Starting of Backing	9074	0.14%
	60 - Off-Highway Vehicle Violation	79	0.00%
	61 - Child Restraint	62	0.00%
	62 - Seat Belt	106	0.00%
	63 - Seat Belt (Equipment)	5	0.00%
MOVEMENT:	Not Stated	110959	1.66%
Movement	A - Stopped	937251	14.04%

Preceding Collision	B – Proceeding Straight	3037926	45.49%
	C – Ran Off Road	184331	2.76%
	D – Making Right Turn	199623	2.99%
	E – Making Left Turn	505106	7.56%
	F – Making U-Turn	46116	0.69%
	G – Backing	107350	1.61%
	H – Slowing/Stopping	366046	5.48%
	I – Passing Other Vehicle	40422	0.61%
	J – Changing Lanes	245857	3.68%
	K – Parking Maneuver	15157	0.23%
	L – Entering Traffic	97336	1.46%
	M – Other Unsafe Turning	121627	1.82%
	N – Crossed Into Opposing Lane	32600	0.49%
	O – Parked	407489	6.10%
	P – Merging	21938	0.33%
	Q – Traveling Wrong Way	15521	0.23%
	R – Other	184871	2.77%
INATTN: Inattention	Not Stated	6612137	99.02%
	A – Cell Phone Handheld	6757	0.10%
	B – Cell Phone Hands Free	638	0.01%
	C – Electronic Equip	1889	0.03%
	D – Radio/CD	5998	0.09%
	E – Smoking	916	0.01%
	F – Eating	2518	0.04%
	G – Children	2877	0.04%
	H – Animal	985	0.01%
	I – Personal Hygiene	362	0.01%
	J – Reading	1224	0.02%
	K – Other	40561	0.61%
	P – Cell Phone	664	0.01%
Overall		6677526	100%

② 인간의 특성에 따라 분류된 데이터 항목 및 분포

Data Field	Values	Party Count	Proportion, %
PTYPE: Party Type	Not Stated	10462	0.16%
	1 - Driver (including Hit and Run)	6023108	90.20%
	2 - Pedestrian	104456	1.56%
	3 - Parked Vehicle	426983	6.39%
	4 - Bicyclist	89481	1.34%
	5 - Other	23036	0.34%
PSEX: Party Sex	Not Stated	869100	13.02%
	F - Female	2269910	33.99%
	M - Male	3538516	52.99%
PAGE: Party Age	Not Stated	1026205	15.37%
	under 18	420324	6.29%
	19-30	1830456	27.41%
	31-40	1115388	16.70%
	41-50	1041339	15.59%
	51-64	866588	12.98%
	over 65	377226	5.65%
INSURED: Financial Responsibility	Not Stated	652929	9.78%
	E - Used if the officer is called away from the scene of the collision prior to obtaining the insurance information	731	0.01%
	N - No Proof of Insurance Obtained	540126	8.09%
	O - Not Applicable (used for parked cars, bicyclists, pedestrians, and party type others)	643286	9.63%
	Y - Yes, Proof of Insurance Obtained	4840454	72.49%
PRACE: Party Race	Not Stated	1281613	19.19%
	A - Asian	420889	6.30%
	B - Black	460885	6.90%

	H - Hispanic	1765627	26.44%
	O - Other	323888	4.85%
	W - White	2424624	36.31%
Overall		6677526	100%

③ 차량 특성에 따라 분류된 데이터 항목 및 분포

Data Field	Values	Party Count	Proportion, %
SBUS: School Bus Related	Not Stated	6658595	99.72%
	E - School Bus Related	18931	0.28%
VEHYEAR: Vehicle Year	Not Stated	613505	9.19%
	2003	431082	6.46%
	2000	425897	6.38%
	2001	423333	6.34%
	2004	414238	6.20%
	2002	413182	6.19%
	Others	3956289	59.25%
VEHMAKE: Vehicle Make	TOYOTA	956845	14.33%
	FORD	944114	14.14%
	HONDA	701024	10.50%
	CHEVROLET	655090	9.81%
	Others	3420453	51.22%
VEHTYPE: Statewide Vehicle Type	Not Stated	1180514	17.68%
	A - Passenger Car/Station Wagon	4058136	60.77%
	B - Passenger Car with Trailier	14274	0.21%
	C - Motorcycle/Scooter	90261	1.35%
	D - Pickup or Panel Truck	800745	11.99%
	E - Pickup or Panel Truck with Trailer	32476	0.49%
	F - Truck or Truck Tractor	82048	1.23%
	G - Truck or Truck Tractor with Trailer	121105	1.81%

	H - Schoolbus	16400	0.25%
	I - Other Bus	27511	0.41%
	J - Emergency Vehicle	38386	0.57%
	K - Highway Construction Equipment	2151	0.03%
	L - Bicycle	88285	1.32%
	M - Other Vehicle	22616	0.34%
	N - Pedestrian	101944	1.53%
	O - Moped	674	0.01%
Overall		6677526	100%

④ 사고 심각도에 따라 분류된 데이터 항목 및 분포

Data Field	Values	Party Count	Proportion, %
PKILLED: Killed Victim Count in Party	0	6653749	99.64%
	1	22220	0.33%
	Over 2	1557	0.02%
PINJURED: Injured Victim Count in Party	0	5159084	77.26%
	1	1277844	19.14%
	2	184005	2.76%
	Over 3	56593	0.85%
Overall		6677526	100%

⑤ 기타 요인에 따라 분류된 데이터 항목 및 분포

Data Field	Values	Party Count	Proportion, %
OAF1: Primary Other Associated Factor	Not Stated	426664	6.39%
	A - Violation	504084	7.55%
	E - Vision Obscurements	41400	0.62%
	F - Inattention	216701	3.25%
	G - Stop and Go Traffic	204247	3.06%
	H - Entering/Leaving Ramp	65640	0.98%

	I - Previous Collision	36528	0.55%
	J - Unfamiliar With Road	11695	0.18%
	K - Defective Vehicle Equipment	15954	0.24%
	L - Uninvolved Vehicle	30497	0.46%
	M - Other	56527	0.85%
	N - None Apparent	5053903	75.69%
	O - Runaway Vehicle	13686	0.20%
OAF2: Secondary Other Associated Factor	Not Stated	6546740	98.04%
	E - Vision Obscurements	2745	0.04%
	F - Inattention	19883	0.30%
	G - Stop and Go Traffic	12682	0.19%
	H - Entering/Leaving Ramp	9171	0.14%
	I - Previous Collision	5198	0.08%
	J - Unfamiliar With Road	2864	0.04%
	K - Defective Vehicle Equipment	2398	0.04%
	L - Uninvolved Vehicle	4307	0.06%
	M - Other	8042	0.12%
	N - None Apparent	62120	0.93%
	O - Runaway Vehicle	1379	0.02%
Overall		6677523	100%

다) PeMS(Caltrans Performance Measurement System)

- (1) 캘리포니아 주정부 교통국 (CALTRANS)에서 운영하는 시스템으로 캘리포니아의 실시간 교통흐름상황 데이터가 기록된 대용량 데이터베이스
- (2) 30초 간격으로 데이터를 수집하는 루프 감지기에서 데이터를 받아들여 저장하고, 교통 흐름의 속도, 도로 점유율 등의 정보를 제공함
- (3) 지정 시간대, 지정 도로별 교통상황의 변화를 측정할 수 있음
- (4) 다음과 같은 정보 수집원으로부터 교통 데이터를 수집
 - (가) 지능형 교통시스템(ITS)의 일부로 고속도로 전역에 걸쳐 설치된 차량검지기 (VDS, Vehicle Detector Stations)
 - 교통량, 도로 점유율, 교통 흐름 속도 등의 데이터를 얻을 수 있음
 - (나) 차량 계수기 (Traffic Counter)
 - (다) 고속도로 순찰대 및 교통국으로부터 수집된 사건 사고 데이터
 - (라) 차선 폐쇄, 포토로그, 통행료 전자지불 정보 등 기타 데이터
 - (마) 다음과 같은 항목들에 대한 기록을 제공함
 - ① Stations: 감지센서 정보, 센서의 위치(구역, 도시, 도로상 거리 기준 위치 등), 이름, 종류, ID 등의 항목 제공
 - ② Spatial Analysis: 센서별로 측정 시각, 교통량(Flow), 점유율(Occupancy), 교통 흐름 속도(Speed) 등의 항목을 제공, 차선별 측정값이 존재하며, 전체 차선에 대한 평균값 또한 제공
 - ③ Aggregates: 일정 시간 동안의 누적 통행량과 관련된 값들을 측정, 총 차량 이동 거리(VMT, Vehicle-Miles Travelled), 총 차량 이동 시간(VHT, Vehicle-Hours Travelled), 지체 시간(Delay) 등의 항목 제공
 - ④ Bottlenecks: 교통체증이 주로 일어나는 병목 구간에 대한 정보를 제공, 병목이 존재하는 도로, 해당 구간 위치, 발생 시간대, 구간 길이, 병목 상태 유지 시간, 평균 지체량 등의 항목 제공
 - ⑤ AADT: 고속도로 구간별 연 평균 일당 통행량 (Annual Average Daily Traffic) 정보 제공
 - ⑥ Level of Service: 서비스 수준(Level of Service)을 A등급에서 F등급까지 나누어 해당 고속도로에서의 한 시간당 서비스 수준 분포를 기록

⑦ Detector Health: 각 검지기들에 대하여 날짜 별로 검지기 위치, 구간 정보, 차선별 센서 상태를 기록

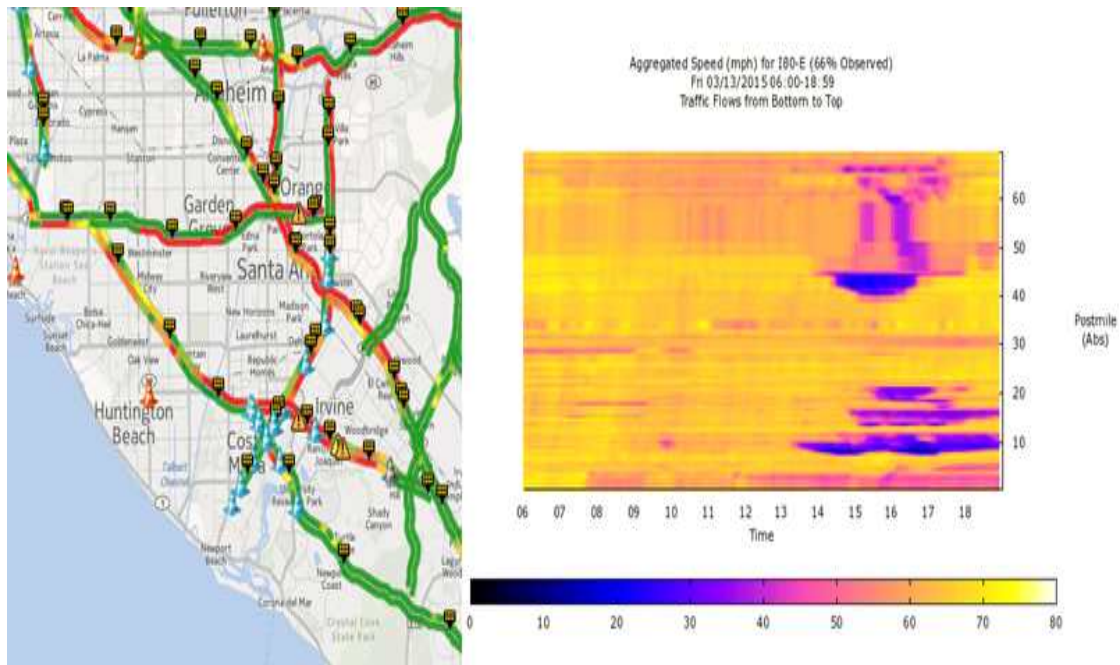


그림 9 PeMS 제공 실시간 교통정보 데이터 (좌)실시간 교통상황 (우) 일간 교통상황 시각화
출처: Caltrans PeMS website (<http://pems.dot.ca.gov/>)

라) 도로교통공단 교통사고 자료 (TAAS)

- (1) 도로교통공단에서는 경찰에서 조사 및 처리한 교통사고 정보를 통합한 데이터를 제공함
- (2) 데이터는 크게 사고정보 데이터와 사상자정보 데이터가 존재함
- (3) 사고 데이터베이스의 경우 사고 발생 날짜 및 발생 시간대, 발생위치, 사고유형, 도로종류 및 형태, 사고당시 기상상태 및 노면상태, 신호기 및 중앙분리시설 유무, 교통장애유무, 당사자정보 (성별, 연령, 면허종류, 면허경과년도, 음주운전 유무) 및 차량정보 (차량용도, 차종), 법규위반정보, 행동유형, 사상자정보 등을 제공
- (4) 사상자 데이터베이스의 경우 사고에 포함된 모든 인원들에 대한 당사자 구분, 성별, 연령, 사고 시 상태, 보호 장구 착용 여부 등의 정보를 제공

년도	발생건수(건)		사망자수(명)		부상자수(명)	
	건	전년대비 증감률	명	전년대비 증감률	명	전년대비 증감률
1995	248,865	-	10,323	-	331,747	-
1996	265,052	6.50%	12,653	22.50%	355,962	7.20%
1997	246,452	-7.10%	11,603	-8.30%	343,159	-3.60%
1998	239,721	-2.80%	9,057	-22.00%	340,564	-0.80%
1999	275,938	15.10%	9,353	3.20%	402,967	18.30%
2000	290,481	5.20%	10,236	9.40%	426,984	5.90%
2001	260,579	-10.30%	8,097	-20.90%	386,539	-9.50%
2002	231,026	-11.40%	7,222	-10.90%	348,149	-10.00%
2003	240,832	4.20%	7,212	-0.20%	376,503	8.10%
2004	220,755	-8.40%	6,563	-9.00%	346,987	-7.90%
2005	214,171	-3.00%	6,376	-2.90%	342,233	-1.40%
2006	213,745	-0.20%	6,327	-0.80%	340,229	-0.60%
2007	211,662	-1.00%	6,166	-2.60%	335,906	-1.30%
2008	215,822	1.90%	5,870	-4.90%	338,962	0.90%
2009	231,990	7.40%	5,838	-0.60%	361,875	6.70%
2010	226,878	-2.30%	5,505	-5.80%	352,458	-2.70%
2011	221,711	-2.30%	5,229	-5.10%	341,391	-3.20%
2012	223,656	0.80%	5,392	3.10%	344,565	0.90%
2013	215,354	-3.80%	5,092	-5.60%	328,711	-4.70%
2014	223,552	3.80%	4,762	-6.50%	337,497	2.60%
2015	232,035	3.70%	4,621	-3.00%	350,400	3.80%

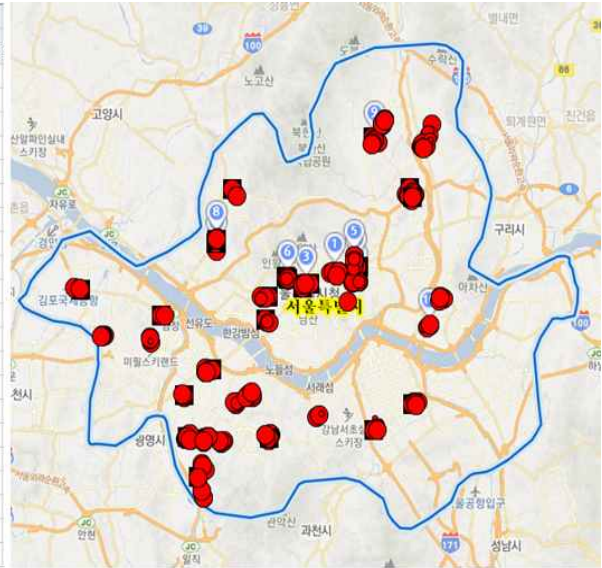


그림 10 TAAS 제공 (좌)교통사고 추세 (우) 사고다발지역 시각화

출처: TAAS website (taas.koroad.or.kr)

마) 서울 속도 데이터

- (1) 서울시 자체적으로 영업용 택시 (31,300대)의 위치정보 가공 알고리즘을 개발하여 1,430km 구간에 대한 통행속도 분석 프로세스를 구축함
- (2) 카드 단말기에 GPS 운행기록을 10초 단위로 저장 후 매 2분 30초마다 수집하여 5분 단위로 통행속도를 생성함
- (3) 도로기능/도로유형/지역/교량, 월/요일/시간대 (오전/낮/오후) 별 통행속도 데이터가 서울시 교통정보센터 웹페이지에 보고서 형태로 공개되어 있음

바) 구급 및 응급 데이터

- (1) 병원 전 단계의 응급의료서비스체계(Emergency Medical Service System)인 119 구급활동의 기초통계 자료로 구급일지 데이터베이스가 있음
- (2) 구급일지 데이터베이스에는 119 신고접수, 구급차 출동, 현장 도착, 현장출발 및 병원도착 시각 데이터와 사고발생지점의 GPS 정보, 사고유형 및 환자상태, 구급처치와 이송병원 정보가 기록되어 있음
- (3) 병원단계의 응급의료서비스체계에 관한 자료는 국가응급진료정보망(National Emergency Medical Center) 데이터베이스에 기록되어 있음

- (4) 응급진료정보망 데이터베이스에서는 병원 응급실 내원환자의 내원시각, 내원경로 및 수단, 증상 및 치료, 퇴원시각, 그리고 교통사고 보호장구 유무여부를 알 수 있음
- (5) 사고 유형을 바탕으로 구급일지 데이터베이스와 응급진료정보망 데이터베이스를 통합, 분석할 시 교통사고 환자마다 병원 전 단계 및 병원 단계에서 소요한 시간, 받은 조치의 유형, 그리고 상해 정도에 대해 파악할 수 있음

5) 통합 데이터 선정

가) Caltrans Traffic Safety Investigation의 Table C를 통해 나타난 사고다발지역에 SWITRS 데이터베이스에 기록된 사고 위치에 대한 정보를 이용하여 사고다발지역에 상세정보를 알아낼 수 있음

나) SWITRS를 이용한 사고다발지역 분석

- (1) SWITRS를 활용하여 아래와 같이 Table C로부터 도출된 사고 다발지역 (사각형으로 표기된 부분)에서 고속도로상의 기타 구간에 비해 사고가 빈번히 일어남을 실제로 확인할 수 있음

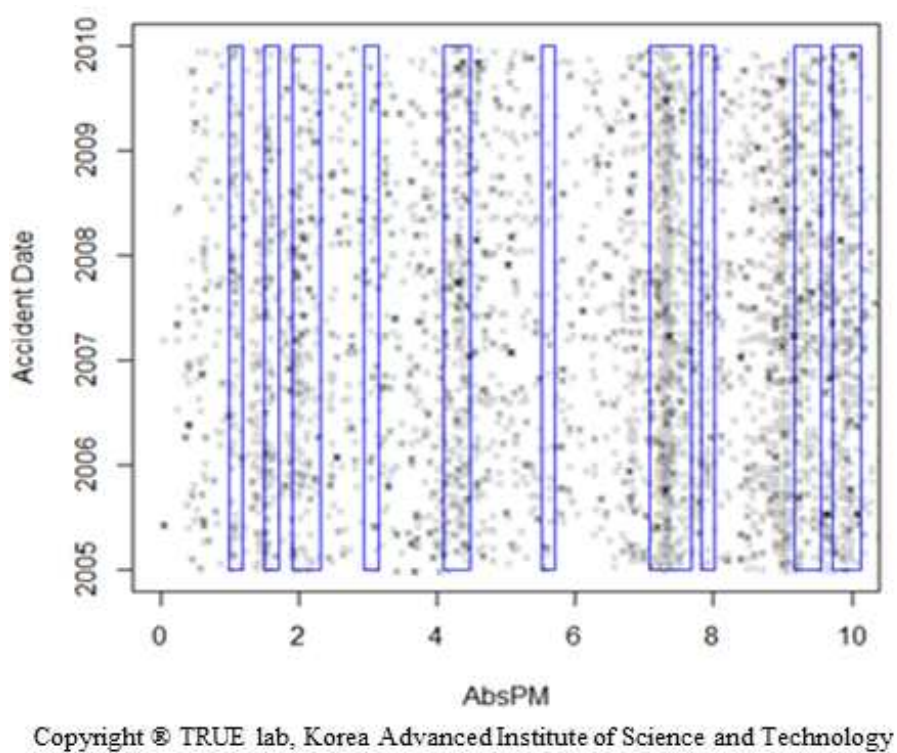
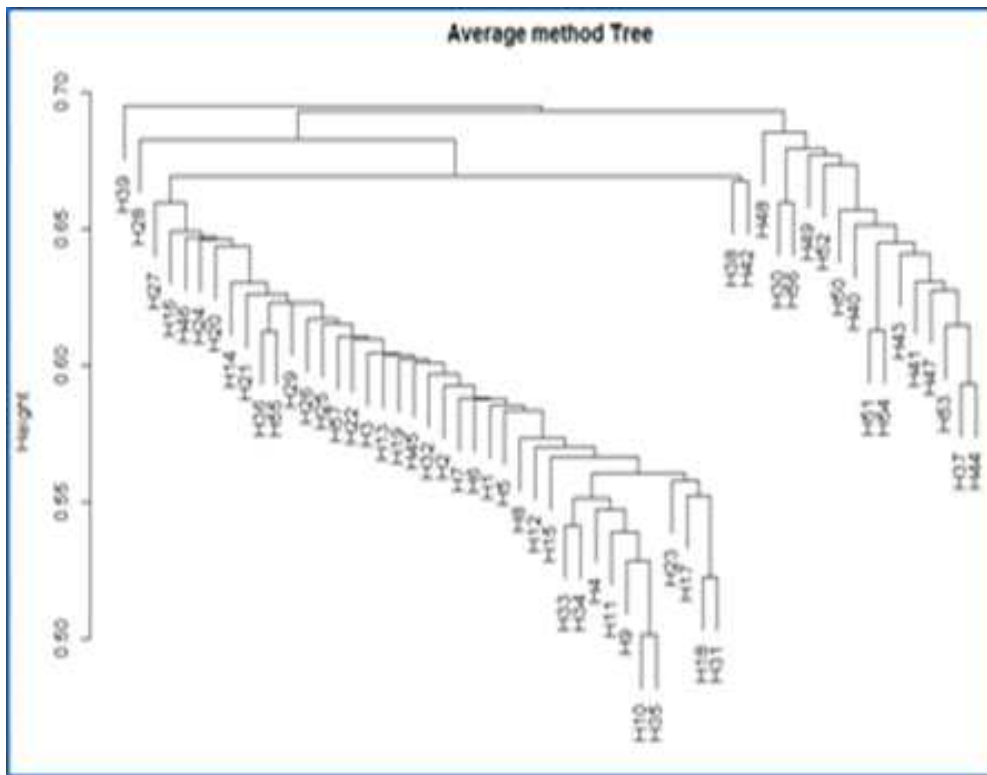


그림 11. SWITRS를 이용한 사고다발지역 분석 결과

- (2) SWITRS에 기록된 전체 사고 데이터베이스 중에서 사고다발지역에서 일어난 사고 정보들을 추출하여, 사고다발지역 별로 해당 지역 내에서 일어난 사고들의 특징을 분류할 수 있음

(가) 사고다발지역에서 일어난 사고들에 대한 위계적 군집분석 (Hierarchical Clustering) 분석 결과, 아래와 같이 사고다발지역들이 분류될 수 있음을 확인



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

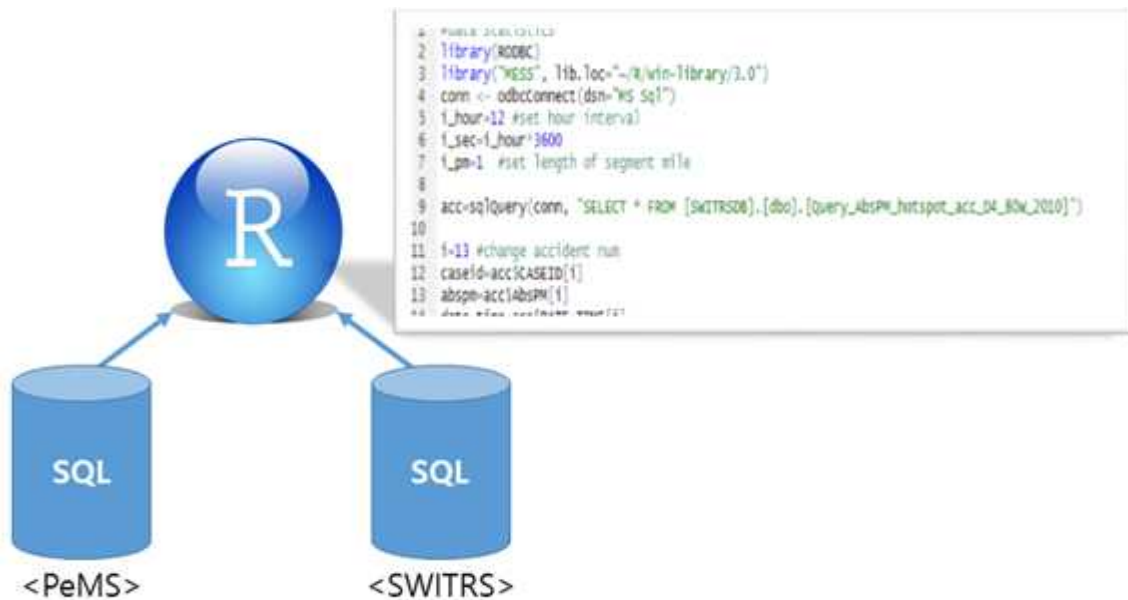
그림 12. 사고다발지역 사고들에 대한 위계적 군집분석 결과

- 다) Table C와 SWITRS를 이용하여 사고다발지역에서 일어난 사고에 대한 정보를 얻을 수 있지만, 사고가 일어난 시점에서 사고지점 주변의 교통 흐름에 대한 정보는 얻을 수 없음
- 라) 사고가 일어난 시점에서 사고지점 주변의 교통흐름에 대한 정보를 얻기 위해 캘리포니아 교통국에서 운영하는 PeMS (Performance Measurement System)에서 교통 흐름 데이터를 활용할 수 있음
- 마) PeMS는 실시간으로 캘리포니아 전역의 고속도로에 설치된 39,000개 이상의 센서로부터 교통 흐름과 관련된 데이터를 수집하고 저장함
- 바) 교통흐름센서는 교통류 속도 (flow speed), 점유율 (occupancy) 등을 실시간으로 측정함
- 사) 사고다발지역에서 일어난 사고 전후의 교통흐름을 분석하기 위해, 5분 간격으로 사고지역 주변의 교통류 속도를 기록한 데이터를 이용함

나. 위험요인 분석을 위한 통합데이터 플랫폼 개발

1) 데이터 재구성을 통한 데이터 전처리 (preprocessing) 과정 수행

가) 데이터 통합을 위하여 SWITRS (교통사고데이터)와 PeMS (교통흐름데이터)를 SQL 형태로 데이터베이스화해두고, 쿼리하여 통합



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 13. 데이터베이스 통합 프로세스

- (1) PeMS 내 30초 마다 생성된 원데이터(Raw data)를 이용하여 5분 데이터를 생성함 (aggregation)
- (2) 각 차선 별 검지기의 상태를 고려하여 만약 3초 데이터에서 해당되는 데이터가 비어있을 경우 (검지기의 문제 또는 교통량 0) imputed value를 사용 (regression, pattern analysis 등을 이용)하고, 대신 observed에 %로 관측된 샘플의 비율을 표시, 즉 100%는 완벽한 30초 데이터를 모두 사용한 5분 데이터를 의미함
- (3) SWITRS와 PeMS의 5분 데이터를 데이터베이스화해두고, 사고위치(post mile)와 사고 시각을 기준으로 필요에 따라 데이터베이스로부터 매칭과정을 거쳐 쿼리하여 사용하는 모델로 구현
- (4) 매칭과정 시 주요사항으로 사고 위치와 검지기 위치와의 거리에 대한 기준이 필요하며, 사고 시간은 같은 형식의 데이터 형태를 취하게끔 전처리가 필요하며, 사고위치(postmile) 또한 두 데이터베이스가 같은 기준 지점을 가지도록 조정

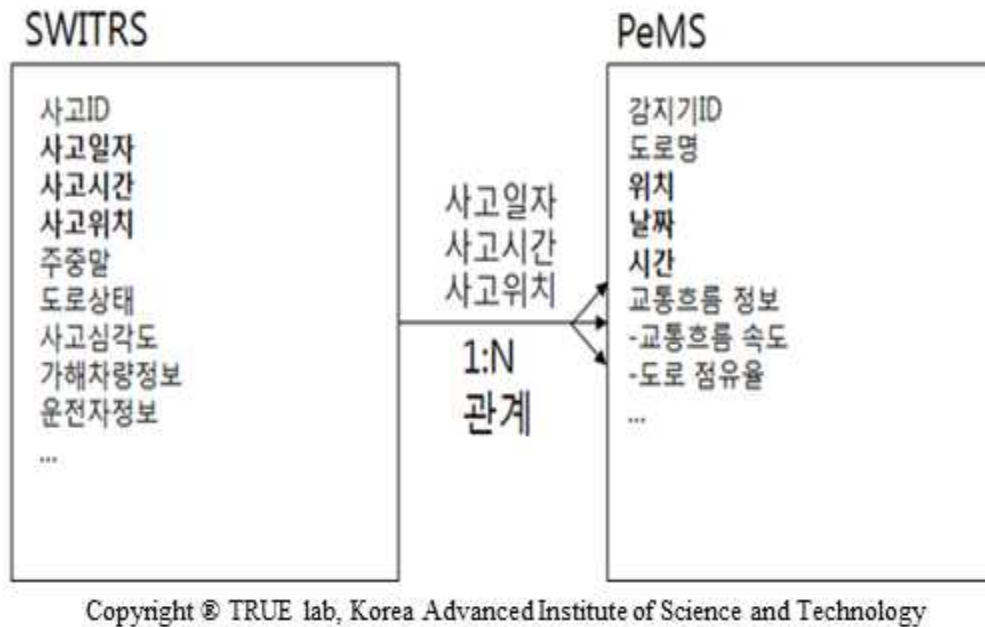


그림 14. PeMS, SWITRS 통합데이터 스키마

다. 데이터 통합 기술 검증

1) 실제 PeMS와 SWITRS 데이터에서 샘플링을 통한 검증 실시

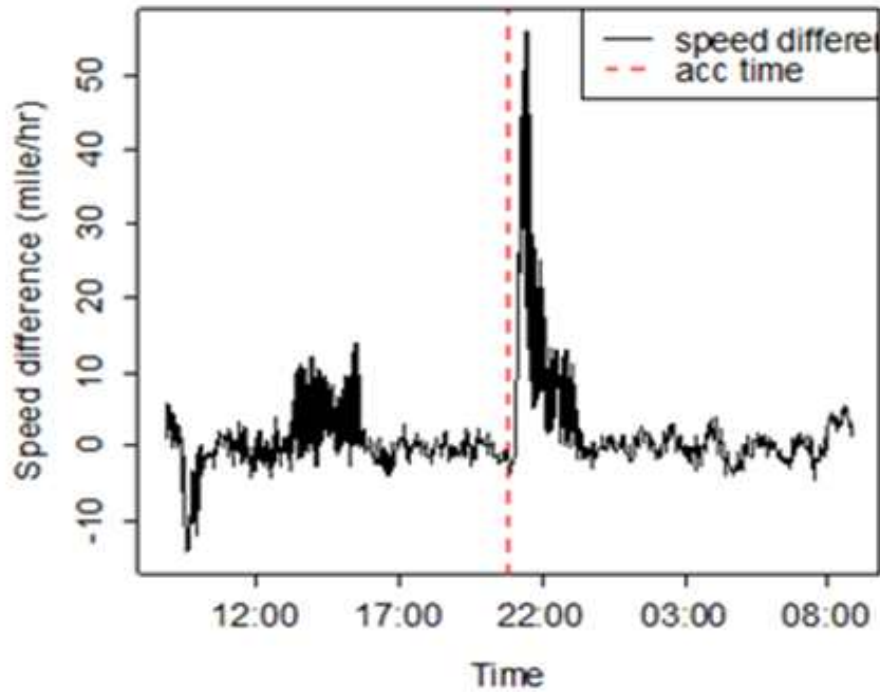
가) 캘리포니아 구역4의 2010년 I-80 West 도로의 데이터를 사용하여 약 11,000,000 개의 PeMS 데이터를 기록 (속도, 점유율, 교통량)

나) SWITRS에서 여기에 해당되는 사고 1,631개와 PeMS 5분 데이터를 데이터베이스화 해두고, 사고위치(post mile)와 사고 시각을 기준으로 필요에 따라 데이터베이스로부터 매칭과정을 거쳐 나온 결과를 분석함

다) 두 개의 데이터베이스를 통합한 데이터 분석 결과 신뢰성 있는 결과를 획득

(1) SWITRS에 기록된 사고지점 정보를 이용하여, PeMS에 기록된 사고가 일어난 지점의 교통 흐름 정보가 사고 시점을 전후로 어떻게 변화하는지 측정할 수 있음

(가) 2010년 3월 25일 20시 51분 경 14.3mile 지점에서 발생한 사고와 PeMS 교통 데이터 매칭 결과 아래와 같이 사고 이후 속도가 급격히 감소하는 현상을 확인할 수 있었으며, 이 속도의 감소가 없었던 평상시 교통 흐름 속도와의 차이를 봤을 때도 그 차이가 상당히 크게 나타남을 확인할 수 있음



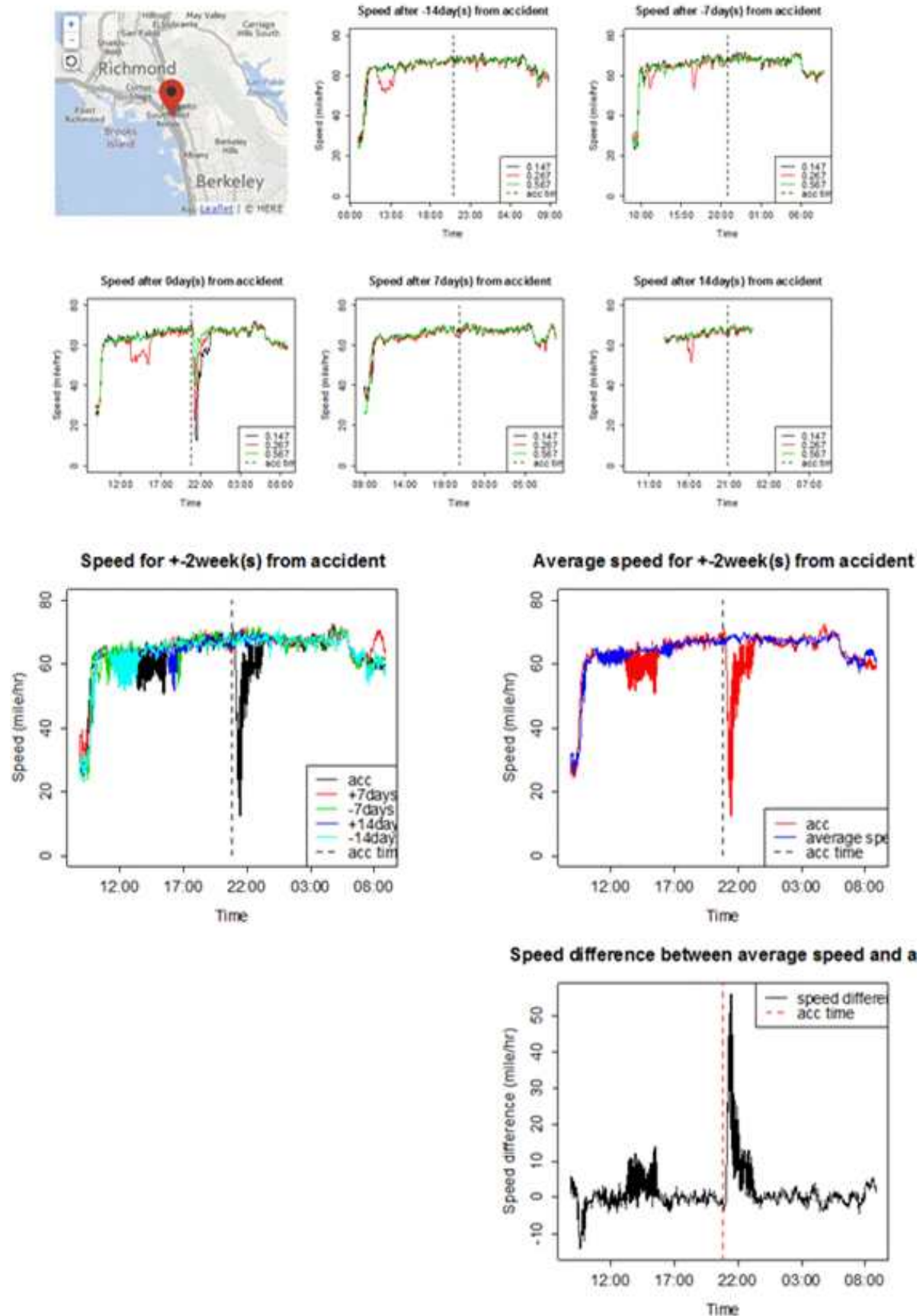
Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 15. 특정 교통사고 시점 이후의 교통 흐름 속도 변화와 평상시 교통 흐름 속도와의 차이

(나) 4건의 사고를 분석해본 결과 4건 모두 특정 교통사고 시점 이후의 교통 흐름 속도와 평상시 교통 흐름 속도와의 차이가 크게 나타나는 신뢰성 있는 결과를 획득할 수 있었으며 이는 데이터가 유효하게 통합되었음을 의미

- ① 사고 지점에서의 교통 흐름 속도가 어떻게 변화하였는지 사고 2주전, 1주전, 사고 당일, 사고 1주 후, 사고 2주 후의 다섯 가지 날짜에 대해 측정
- ② 사고가 발생한 지점 인근 검지기 3개에서 수집된 데이터를 이용하였음
- ③ 사고 당일의 경우에 대해서, 사고가 일어난 직후 주변 교통 흐름 속도가 대폭 감소한 후 천천히 회복되는 것을 관찰할 수 있음
- ④ 사고 2주전, 1주전, 사고 1주 후, 사고 2주 후의 경우에 대해서는, 사고 당일과 달리 교통 흐름 속도에 변화가 미비함을 관찰할 수 있음
- ⑤ 사고 당일의 속도와 사고 2주전, 1주전, 사고 1주 후, 사고 2주 후의 평균속도를 비교해본 결과 사고 당일의 교통 흐름 속도가 사고 2주전, 1주전, 사고 1주 후, 사고 2주후의 평균속도 보다 변화가 큼을 확인할 수 있음

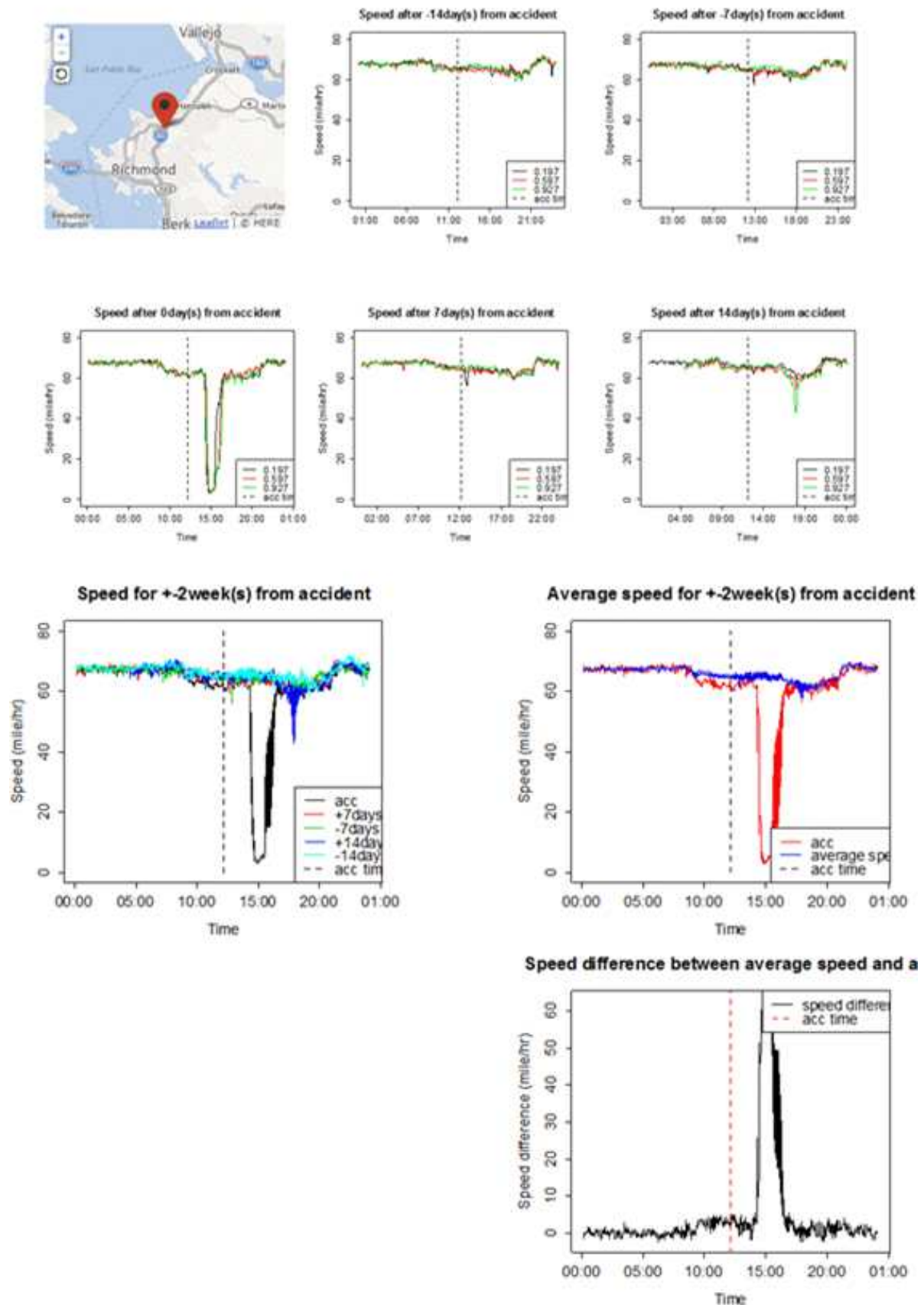
Accident #1: 2010-03-25 20:51:00, 14.306 postmile, not_hotspot, weekday, suburban, fatal, hitobject fixobject, motorcycle



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 16. 사고지점정보와 교통흐름정보를 이용한 분석결과 예시- 1

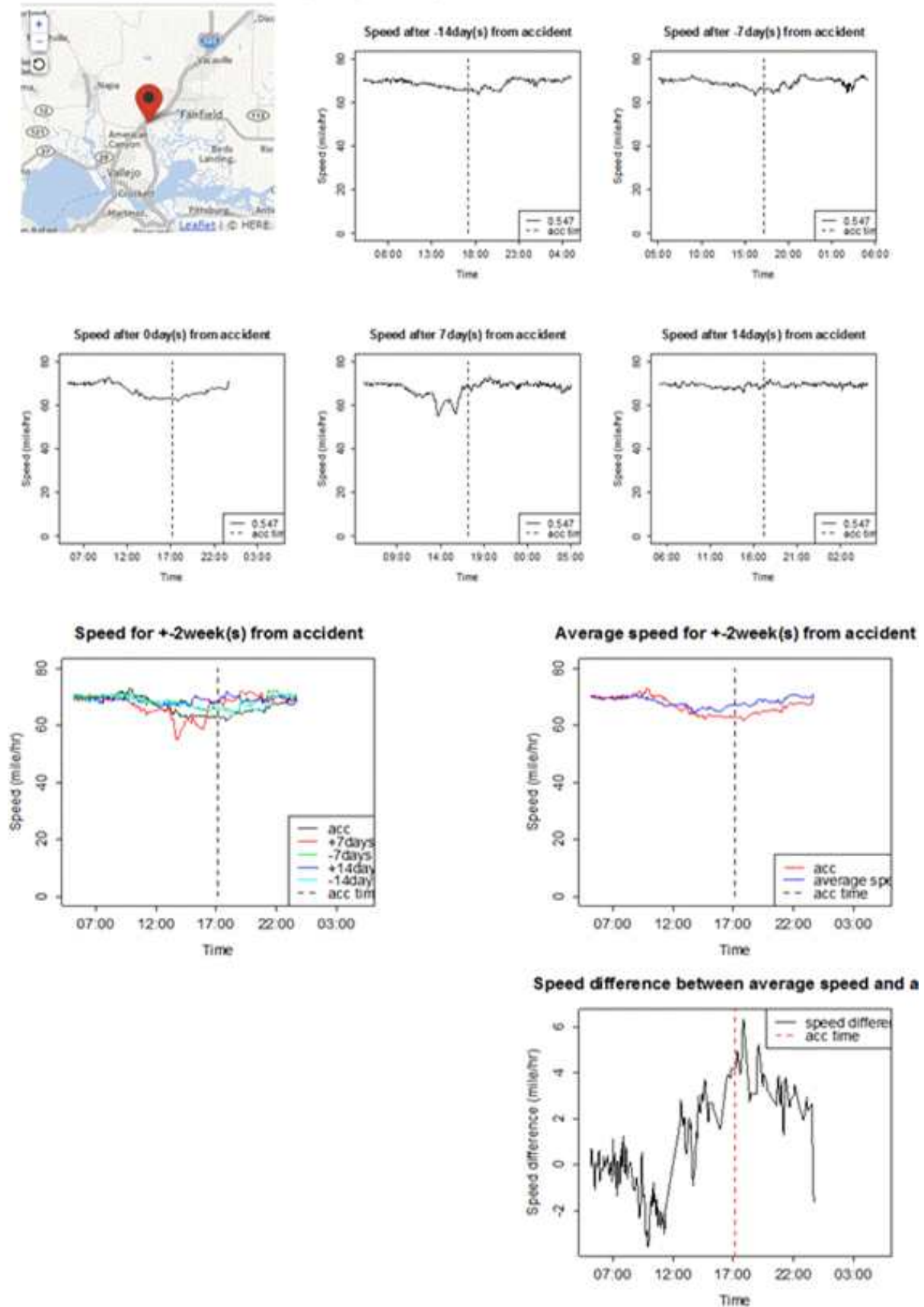
Accident #2: 2010-02-21 12:10:00, 20.026 postmile, not_hotspot, weekend, suburban, level 3 severity, overturned otherobject, passengercar



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 17. 사고지점정보와 교통흐름정보를 이용한 분석결과 예시 - 2

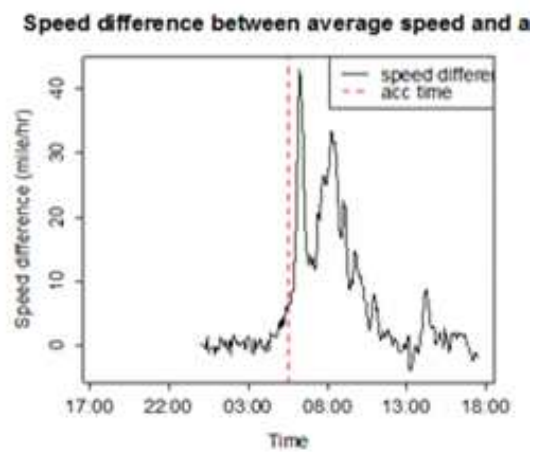
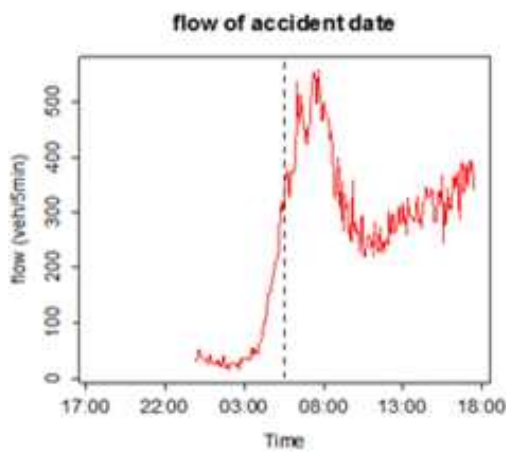
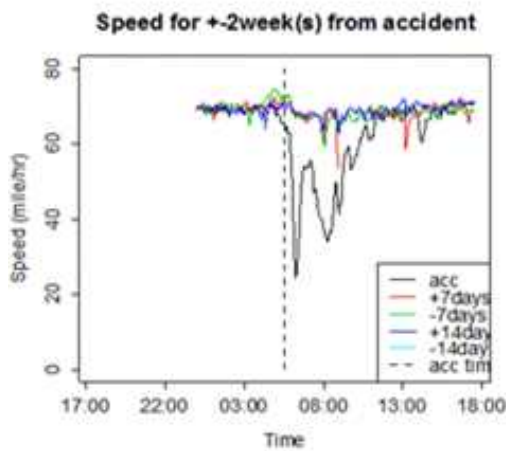
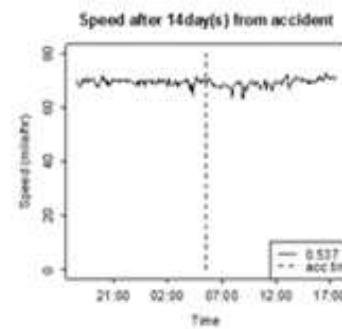
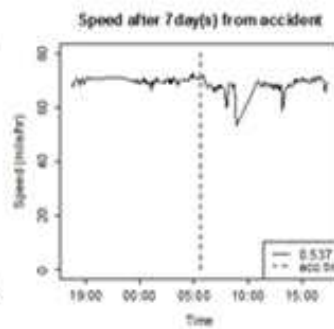
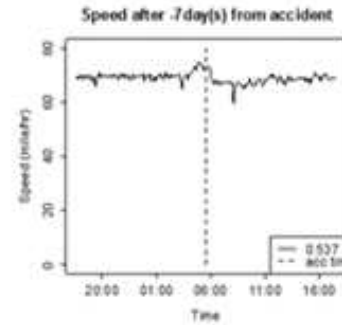
Accident #3: 2010-04-04 17:10:00, 40.926 postmile, not_hotspot, weekend, suburban, PDO severity, hitobject fixobject, passengercar



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 18. 사고지점정보와 교통흐름정보를 이용한 분석결과 예시 - 3

Accident #4: 2010-01-20 05:35:00, 40.936 postmile, not_hotspot, weekday, suburban, PDO severity, hitobject fixobject, passengercar



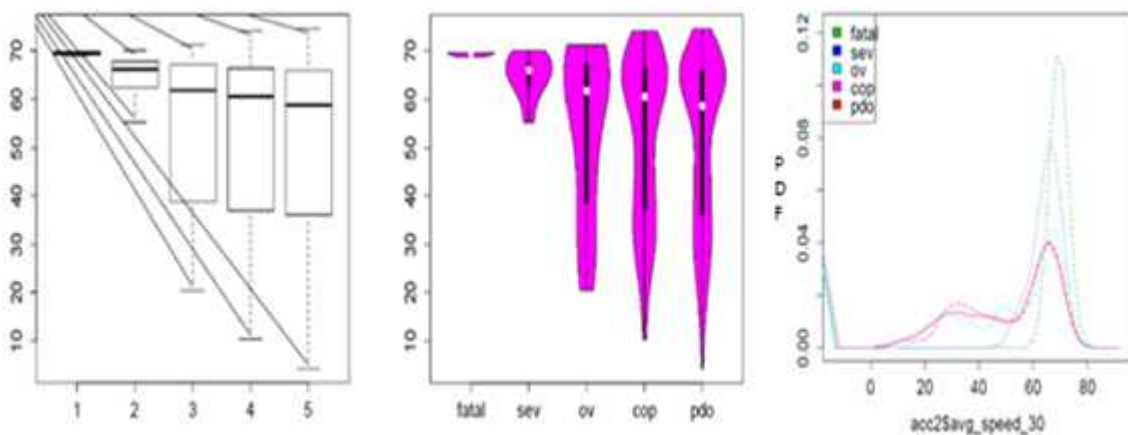
Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 19. 사고지점정보와 교통흐름정보를 이용한 분석결과 예시 - 4

(2) 사고 전 30분간, 사고지점 앞뒤로 0.5 mile 구간의 평균속도를 산정하여 사고 당시의 그 지점의 평균 차량 속도 정보를 이용하여 사고에 미치는 영향 분석

(가) 총 1,600여개 사고 중 그 당시의 평균속도 산정 가능 지역은 약 600개

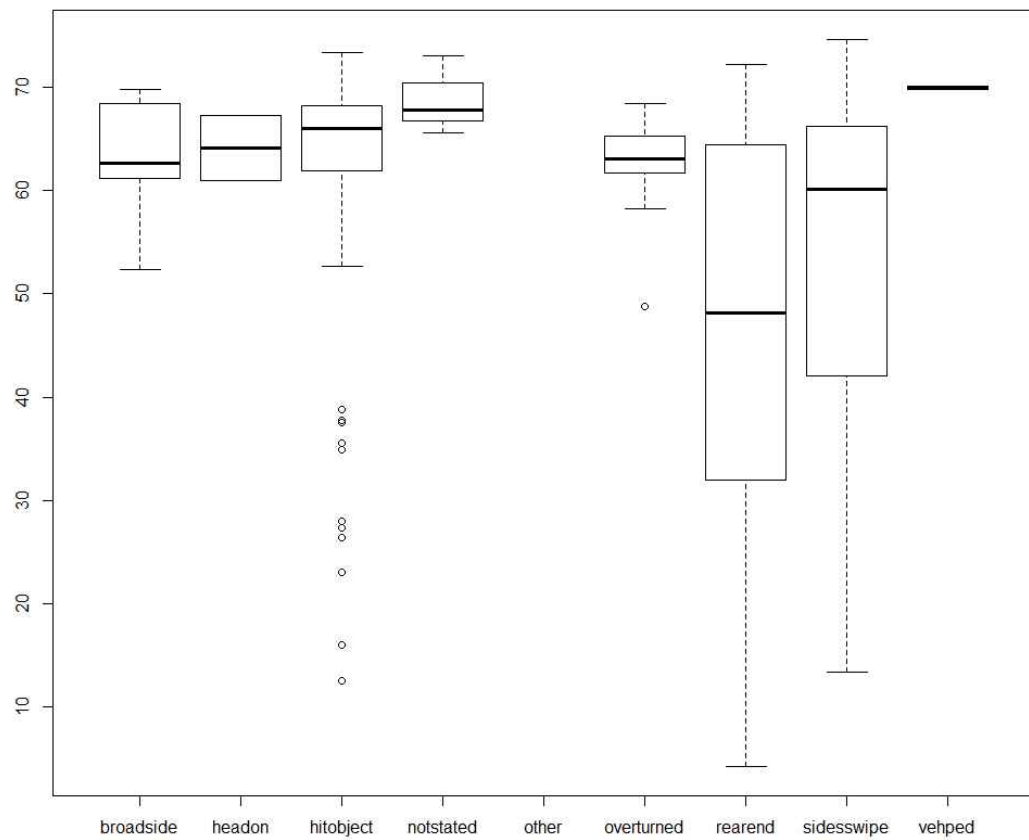
- ① Fatal, Severe, Other Visible, Complaint of pain, Property Damage Only로 분류된 사고 심각도를 기준으로 평균속도를 살펴본 결과, 사고 심각도가 높을수록 평균속도의 중앙값이 높음을 확인할 수 있었음
- ② 사고 심각도가 가장 높은 Fatal과 그 다음으로 높은 Severe의 경우 사고 당시 속도가 사고 심각도가 보다 낮은 Other Visible, Complaint of pain, Property Damage Only와 비교하여 밀도 있게 분포함을 알 수 있음



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 20. 사고 심각도에 따른 평균속도

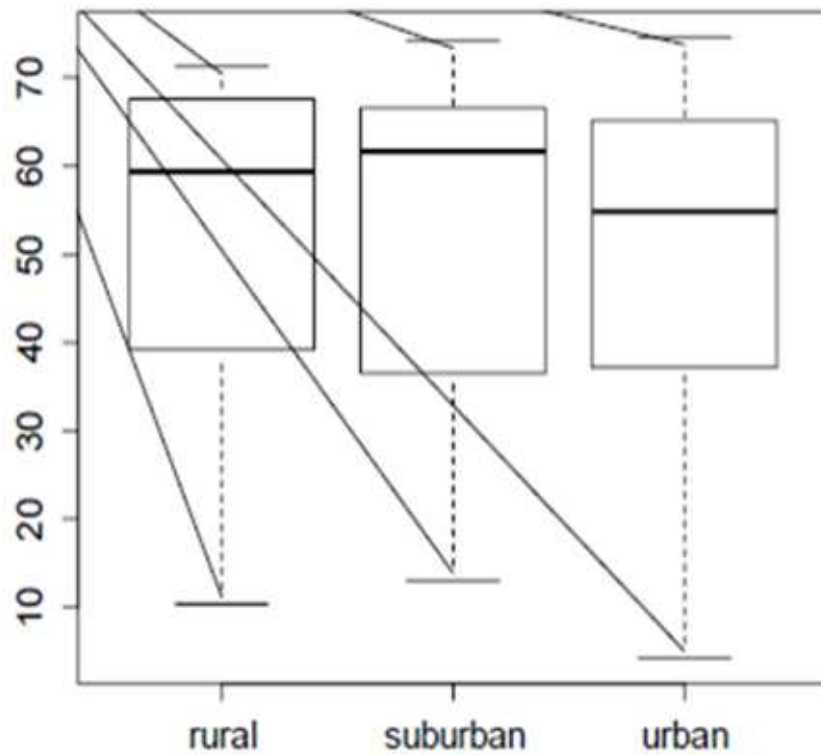
- ③ Head-On, Sideswipe, Rear End, Broadside, Hit Object, Overturned, Vehicle/Pedestrian, Other, Not Stated로 분류된 사고 충돌유형을 기준으로 평균속도를 살펴본 결과, Vehicle/Pedestrian과 충돌한 사고의 경우 가장 높은 중앙값을 보이고 Hit Object가 그 다음 높은 중앙값 보임을 확인할 수 있었음
- ④ Vehicle/Pedestrian과 충돌한 사고의 경우 다른 사고 충돌유형과 비교하여 가장 높은 속도에서 밀도 있게 분포함을 알 수 있으며 Rear End와 Sideswipe 사고 충돌유형의 경우 낮은 속도에서도 발생하는 경향이 있음을 확인할 수 있었음



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 21 사고 충돌 유형에 따른 평균속도

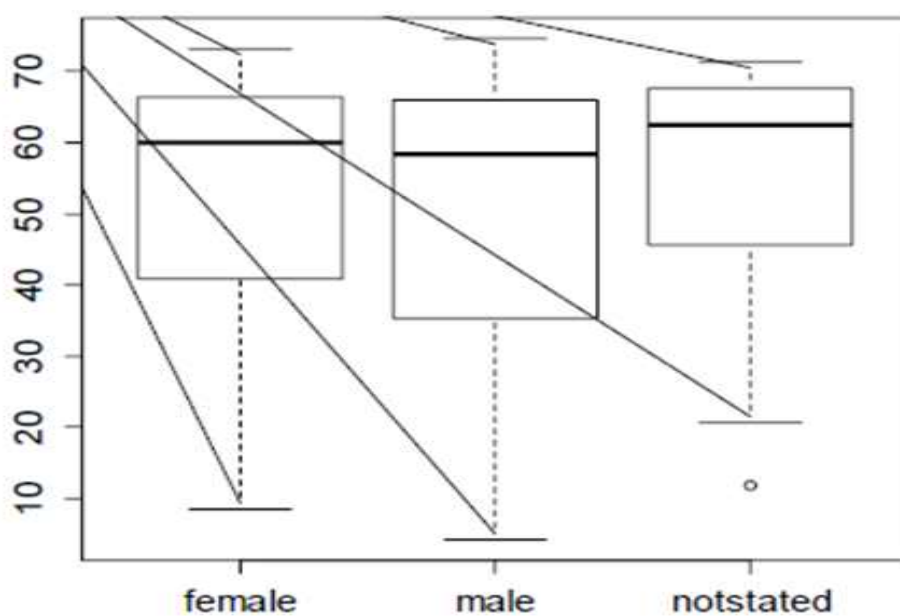
- ⑤ 사고 위치에 따른 평균속도를 살펴본 결과 suburban, rural, urban 순서로 중앙값이 나타남을 확인할 수 있었음
- ⑥ 사고 위치가 평균속도에 분포에 미치는 영향이 미비함을 확인할 수 있었음



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 22. 사고 위치에 따른 평균속도

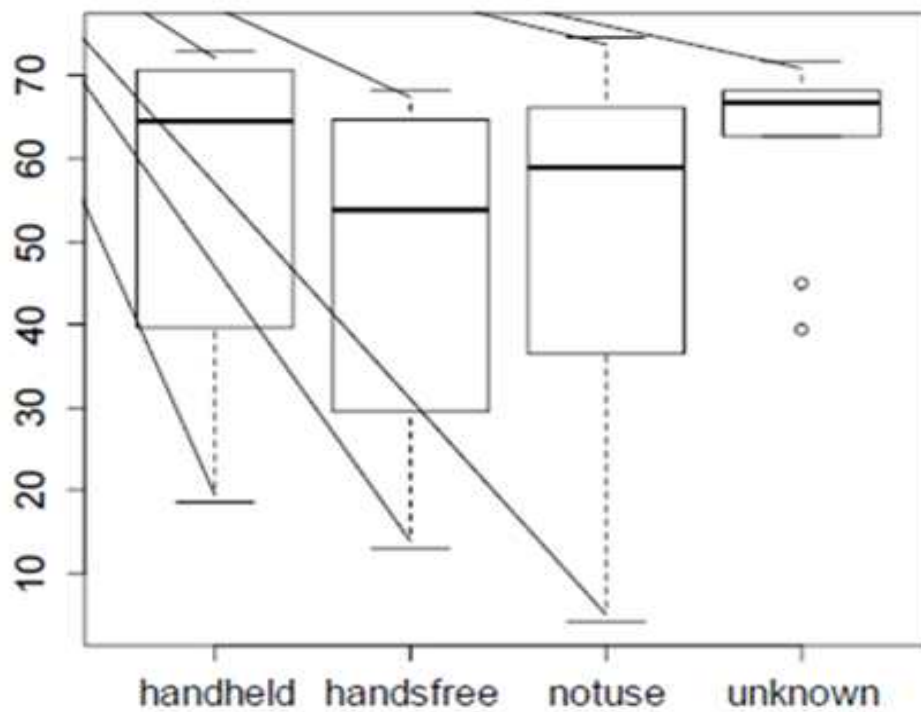
- ⑦ 잘못된 운전자의 성별에 따른 평균속도를 살펴본 결과 여성 운전자가 사고를 일으킨 경우, 남성 운전자가 사고를 일으킨 경우와 비교하여 평균속도에 중앙값이 높으나 그 차이가 미비함을 확인할 수 있었음



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 23. 잘못된 운전자의 성별에 따른 평균속도

- ⑧ 잘못된 운전자의 휴대폰 사용에 따른 평균속도를 살펴본 결과 휴대폰을 손에 들고 있는 경우 발생한 사고의 평균속도가 가장 높은 중앙값을 보이는 것을 확인할 수 있었음



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 24. 잘못된 운전자의 휴대폰사용에 따른 평균속도

2. 교통사고 심각도 결정 모델 개발

가. 교통사고 심각도 결정 모델 연구 배경

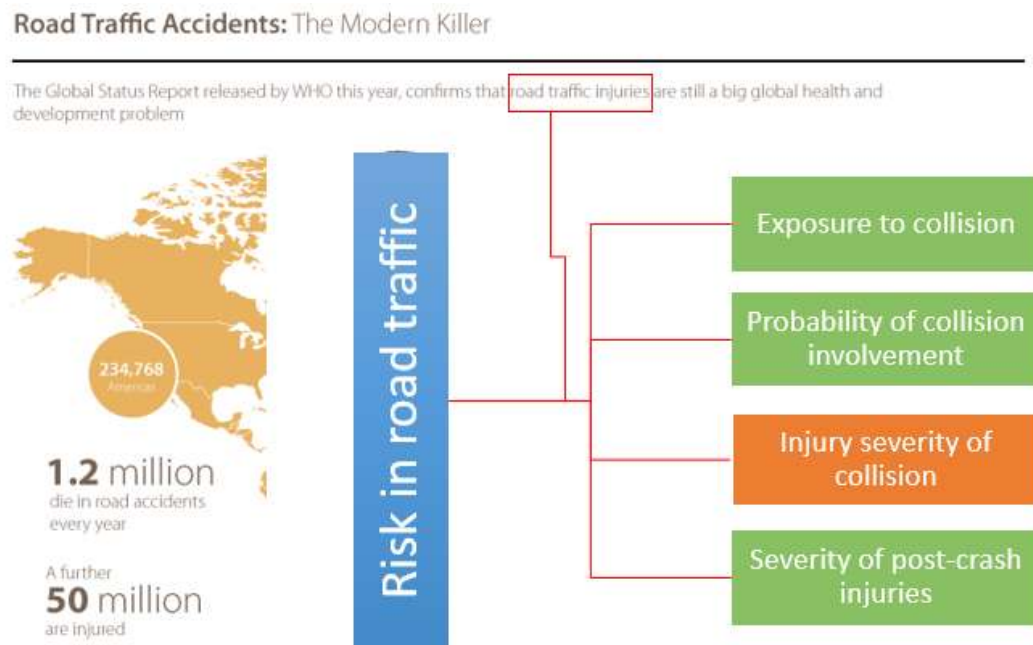


그림 25 도로 교통 위험 요소

출처: Car Claim, Car accident statistics of the world, 2015,

<http://car2claim.blogspot.kr/2015/06/car-accident-statistics-of-world.html>

- 세계 보건기구는 2004년 펴낸 보고서에서 교통사고로 인해 발생하는 피해를 전 지구적 문제점으로 지목함
- 세계보건기구 (WHO)에서 도로 교통에 존재하는 위험, 즉 Risk를 네 가지 요소로 분류하였고 이는 위험에 대한 노출 (exposure to collision), 사고가 일어날 확률 (probability of collision involvement), 사고의 부상 심각도 (injury severity of collision), 사고 후 여파에 의 부상 정도 (severity of post-crash injuries)임
- 본 연구는 앞서 언급된 네 가지 요소 중 사고 부상 심각도에 영향을 주는 요인인 risk factor를 분석하고자 함

나. 연구의 목적

- (1) 대용량 사고 데이터베이스로부터 서로 다른 subgroup, 즉 서로 다른 특징을 가지는 교통사고 연루 차량들의 집단들에 대해 주요 risk factor들이 어떻게 나타나는지 알아보고, 어떤 차이점이 존재하는지를 알아보고자 함
- (2) 본 연구를 수행하기 위해 데이터 마이닝 기법을 활용하였으며 첫 번째로 regularized random forest 기법을 기반으로 여러 subgroup들의 risk factor 순위를 정의함
- (3) 두 번째는 subgroup 관계에 나타나는 risk factor 순위의 변화를 정성적, 정량적으로 분석하여 subgroup간의 비유사도를 평가함

다. 선행 연구 조사 및 본 연구의 차별점

- (1) Risk factor analysis 연구에 대한 선행 연구는 다음과 같은 연구들이 존재하며 로지트 (logit) 및 프로비트 (probit) 회귀분석을 활용한 기존 교통사고 심각도 분석에 관한 선행 연구를 수행함

Author	Subject of analysis	Methodology	Input variables
Shankar et al. (1996)	General accident	Nested logit	Environmental condition, highway design, accident type, etc.
Al-Ghamdi (2002)	General accident	Logistic regression	Accident location and cause of accident
Yan et al. (2005)	Rear-end collision	Logistic regression	Road environmental factors, vehicle type, driver age, alcohol/drug use, driver info
Morgan and Mannering (2011)	General accident	Mixed logit	Age and gender of driver, road surface, weather
Chang and Wang (2006)	General accident	CART	Driver, vehicle, road, environmental characteristics
Kashani and Mohaymany (2011)	Accidents in two-lane, two-way rural roads	CART	Time and place of crash, road type, collision type, road surface, seatbelt, etc.
Montella et al. (2011)	Powered two-wheeler	Classification tree, association rule	Area, road type, lighting, weather, pavement, driver info, PTW type, crash type, etc.
Yau (2004)	Private vehicle, goods vehicle, motorcycle	Stepwise logistic regression	District, human, vehicle, safety, environmental and site factors
Mostafaei et al. (2012)	Moped and motorcycle	Logistic regression	Gender, age, helmet, alcohol level, trip purpose, number of passenger, etc.

그림 26 선행 연구 조사 목록

(2) Logit and Probit models

- Shankar et al. (1996)는 환경적 조건, 고속도로 설계, 사고 유형, 운전자 성향, 사고 심각도에 대한 차량 형태의 영향을 실험하고자 nested logit 모델을 사용함

- Al-Ghamdi (2002)는 사고 심각도의 주요 risk factor가 사고 발생 위치와 사고의 원인임을 확인하기 위해 logistic regression 모델을 사용함
- Yan et al. (2005)의 연구에서도 logistic regression 모델을 사용하여 충돌 사고의 위험에 대한 risk factor의 영향을 관측하였음
- Morgan and Mannering (2011)에서는 mixed logit 모델을 활용하여 운전자의 나이와 성별, 도로 표면 상태 및 영향이 사고 심각도에 미치는 영향에 대해 연구하였음

(3) Decision tree models

- Chang and Wang (2006)의 연구에서는 classification and regression tree (CART) 기법을 사용하여 부상 심각도와 risk factor간의 관계를 설립하고자 하였으며, risk factor로서 운전자와 차량의 특징, 도로와 환경 연관 특징 및 기타 사고와 관련된 변수들을 사용하였음
- 또한, Kashani and Mohaymany (2011)는 CART 기법을 사용하여 지방의 이 차선 도로에서 사고에 대한 부상 심각도는 추월과 안전벨트 미착용이 가장 결정적인 요인인 것으로 기술함
- Montella et al. (2011)의 연구에서는 classification tree를 사용하여 도로 유형, PTW 유형, 충돌 유형, 조명, 커브길 배열 (curve alignment)가 PWT (powered two-wheeler) 충돌의 사고 심각도에 책임 요소임을 확인하였음

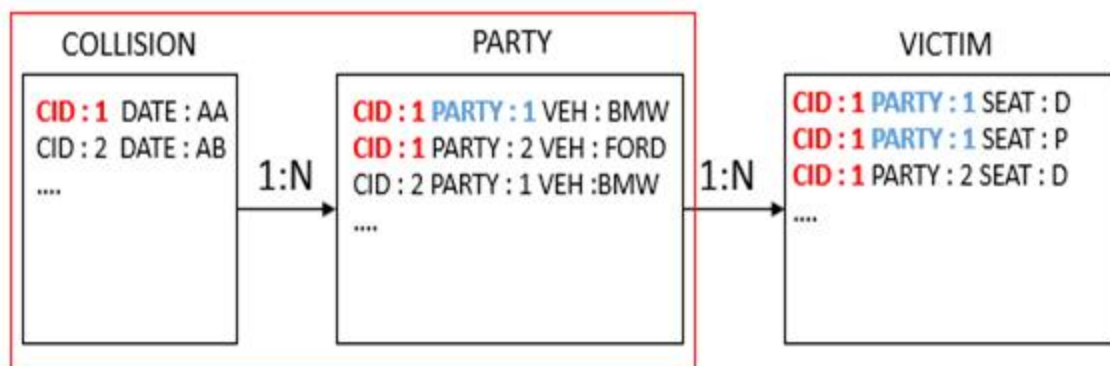
(4) 갈색으로 표시된 것은 단일 집단에 대한 연구이며 파란색으로 표시된 것은 여러 집단에 대한 risk factor analysis를 행한 연구들임

(5) 단일 집단의 연구는 Nested logit, logistic regression 등 각기 다른 방법론 및 데이터, 변수들을 사용하였으므로 직접 비교가 어려우며 복수 집단에 대한 선행 연구의 경우 변수간 상호 의존성을 고려하지 않기에 bias의 가능성이 높은 logistic regression에 기반을 둔 분석 기법이 사용되었으며 주요 risk factor를 찾아내는데 그쳤다는 한계가 존재함

(6) 이에 기존 연구들로부터 차별화를 위해 본 연구에서는 다양한 집단의 risk factor를 하나의 일관된 방식으로 비교하는 것과 분석 과정에서 risk factor 간 의존성을 고려하였음

라. 활용 데이터 및 분류 기준

- 본 연구에 사용된 데이터는 SWITRS 데이터로서 미국 캘리포니아 고속도로 경찰이 수집한 교통사고 데이터로 사고에 대한 전반적 정보, 사고와 연루된 차량 (혹은 보행자) 및 운전자에 대한 정보, 사고와 연루된 피해자들에 대한 정보를 담고 있어 교통사고 위험 요인을 다각적으로 분석할 수 있음
- 데이터 구조는 크게 Collision, Party, Victim의 세 개의 테이블로 구성되며, 본 연구에서는 차량 사고에 연루된 차량의 사고 심각도에 대한 분석을 위함으로 collision과 party 두 테이블의 정보에 대한 2004년부터 2010년까지의 기록을 사용함
- 또한, 데이터에 기록되는 탈 것 항목으로 보행자, 오토바이, 자전거, 트럭, 일반 차량 중 일반 차량에 대한 기록만 사용함



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 27. SWITRS 데이터베이스 구조

- 이렇게 수집된 분석 데이터는 총 2,347,528건의 기록으로 데이터베이스 상에서 5개의 심각도로 구분이 되어 있으며 해당 항목은 사망사고 (fatal), 심각한 부상 (Severe), 육안으로 보이는 부상 (Other visible), 고통 호소 (Complaint of pain), 재물 피해만 존재하는 경우 (Property damage only)임
- 본 연구에서는 부상 정도의 모호함을 해소하고 통계 모델의 정확도 향상을 위해 그림에서 보이는 바와 같이 사망사고 (Fatal), 부상사고 (Injured), 재물 피해 사고 (PDO)의 세 가지 수준으로 재조정함

ALL (2347528 records, 100%)	Fatal (0.27%)	Fatal (0.27%)
	Severe (0.83%)	Injured (25.73%)
	Other Visible (6.49%)	
	Complaint of pain (18.41%)	
	Property Damage Only (74.00%)	PDO (74.00%)

그림 28 사고 심각도 분류

마. Risk Factor 및 Subgroup 정의

(1) Risk Factor

- 동일한 데이터베이스를 사용한 연구인 Kwon 외 저자들의 연구 ‘Application of classification algorithms for analysis of road safety risk factor dependencies’에 의해 다수의 사고 관련 변수들 중 소수의 일부 변수들만이 심각도에 영향을 미치는 위험 요인으로 작용함을 분석하였으며 이를 기반으로 본 연구에서는 SWITRS에서 제공하는 100여개의 교통사고 관련 변수들 중 다음의 11개 변수를 추출하고 risk factor로 지정함

Risk Factor	Description
CRASHTYP	Type of collision
ATFAULT	Whether the party was at fault in the collision
POP	Population
STATEHW	State highway indicator
MOVEMENT	Movement preceding collision
VIOLCAT	Violation category
PSEX	Gender of party
PRACE	Race of party
VEHAGE_Group	Age group of vehicle
PAGE_Group	Age group of party
PSOBER	Party sobriety

그림 29 Risk Factor 선정

- Risk factor로 사용된 11개의 변수는 각각 충돌 유형 (CRASHTYP), 가해/피해차량 여부 (ATFAULT), 충돌지역의 인구 (POP), 주 고속도로 여부 (STATEHW), 충돌 이전 차량 행태 (MOVEMENT), 범규 위반 항목 (VIOLCAT), 운전자 성별 (PSEX), 운전자 인종 (PRACE), 차량 연식 그룹 (VEHAGE_Group), 운전자 나이대 (PAGE_Group), 음주운전 여부 (PSOBER)임

- Risk Factor Description

Risk Factor	Description	Possible Values
CRASHTYP	Type of collision	A - Head-On B - Sideswipe C - Rear End D - Broadside E - Hit Object F - Overturned G - Vehicle/Pedestrian H - Other - - Not Stated
ATFAULT	Whether the party was at fault in the accident	Y - Yes N - No
POP	Population	1 - Incorporated (less than 2500) 2 - Incorporated (2500 - 10000) 3 - Incorporated (10000 - 25000) 4 - Incorporated (25000 - 50000) 5 - Incorporated (50000 - 100000) 6 - Incorporated (100000 - 250000) 7 - Incorporated (over 250000) 9 - Unincorporated (Rural) 0 - University (Private Property) - - Not Stated
STATEHW	State highway indicator	Y - State Highway N - Not State Highway Blank - Not stated
MOVEMENT	Movement preceding	A - Stopped

	collision	B - Proceeding Straight C - Ran Off Road D - Making Right Turn E - Making Left Turn F - Making U-Turn G - Backing H - Slowing/Stopping I - Passing Other Vehicle J - Changing Lanes K - Parking Maneuver L - Entering Traffic M - Other Unsafe Turning N - Crossed Into Opposing Lane O - Parked P - Merging Q - Traveling Wrong Way R - Other - - Not Stated
VIOLCAT	Violation category	01 - Driving or Bicycling Under the Influence of Alcohol or Drug 02 - Impeding Traffic 03 - Unsafe Speed 04 - Following Too Closely 05 - Wrong Side of Road 06 - Improper Passing 07 - Unsafe Lane Change 08 - Improper Turning 09 - Automobile Right of Way 10 - Pedestrian Right of Way 11 - Pedestrian Violation 12 - Traffic Signals and Signs 13 - Hazardous Parking 14 - Lights 15 - Brakes

		16 - Other Equipment 17 - Other Hazardous Violation 18 - Other Than Driver (or Pedestrian) 19 - 20 - 21 - Unsafe Starting or Backing 22 - Other Improper Driving 23 - Pedestrian or "Other" Under the Influence of Alcohol or Drug 24 - Fell Asleep 00 - Unknown - - Not Stated
PSEX	Gender of party	M - Male F - Female - - Not Stated
PRACE	Race of party	A - Asian O - Other B - Black W - White H - Hispanic Blank - Not stated
VEHAGE_Group	Age group of vehicle	~5 years 6~10 years 11~15 years 16~20 years 21~30 years 31~ years
PAGE_Group	Age group of party	~18 19~30 31~40 41~50 51~64 65~
PSOBER	Party sobriety	A - Had Not Been Drinking B - Had Been Drinking, Under

		Influence C - Had Been Drinking, Not Under Influence D - Had Been Drinking, Impairment Unknown G - Impairment Unknown H - Not Applicable - - Not Stated
--	--	--

(2) Subgroup 생성

- Risk factor의 순위를 분석하기 위한 기준으로 14개의 subgroup을 생성하였으며 분류 기준은 인류통계학적 변수 (demographic characteristics), 외부 환경적 변수 (environmental characteristics), 운전 행동 변수 (driving behavior characteristics), 차량 관련 변수 (vehicle characteristics)로 분류함
- 위 네 가지의 특징을 가진 sub group으로서 인류통계학적 변수로서 운전자 성별과 인종이 포함되었으며, 외부 환경적 변수로 인구수, 운전 행동 변수로 가해/피해차량 여부, 차량 관련 변수로 자동차 제조사로 지정함
- 인구수에서 도시 (urban)의 경우 데이터베이스의 인구수 분류 기준에서 가장 높은 인구수 항목인 25만 명 이상에 해당되는 항목이며, 자동차 제조사의 경우 200개가 넘는 항목들 중 일본, 미국, 독일, 한국 각 나라 제조사들 중 가장 많은 기록을 가지고 있는 네 개의 제조사를 선정함

Subgrouping Criteria	# of Subgroups	Subgroup Name	Party count (percentage)	Severity			
				Fatal	Injury	PDO	
Gender of driver	2	Female	1106075 (47.12%)	0.18%	29.02%	70.80%	Demographic characteristics
		Male	1241453 (52.88%)	0.35%	22.79%	76.86%	
Race of driver	4	Asian	224544 (9.57%)	0.15%	22.41%	77.43%	
		Black	185126 (7.89%)	0.24%	29.70%	70.06%	
		Hispanic	660095 (28.12%)	0.36%	26.51%	73.13%	
		White	956381 (40.74%)	0.29%	26.31%	73.40%	
Population	2	Urban	562282 (23.95%)	0.15%	31.25%	68.60%	Environmental characteristic
		Rural	556117 (23.69%)	0.66%	26.48%	72.86%	
At-fault party indicator	2	At-fault	1155204 (49.21%)	0.45%	23.64%	75.91%	Driving Behavior characteristic
		Not-at-fault	1192324 (50.79%)	0.10%	27.75%	72.15%	
Vehicle maker	4	Toyota	392630 (16.73%)	0.19%	26.72%	73.09%	Vehicle characteristic
		Ford	280179 (11.94%)	0.32%	25.56%	74.12%	
		BMW	60954 (2.60%)	0.20%	24.11%	75.69%	
		Hyundai	35239 (1.50%)	0.25%	27.50%	72.25%	

Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 30. Subgroup description

바. 연구 방법론

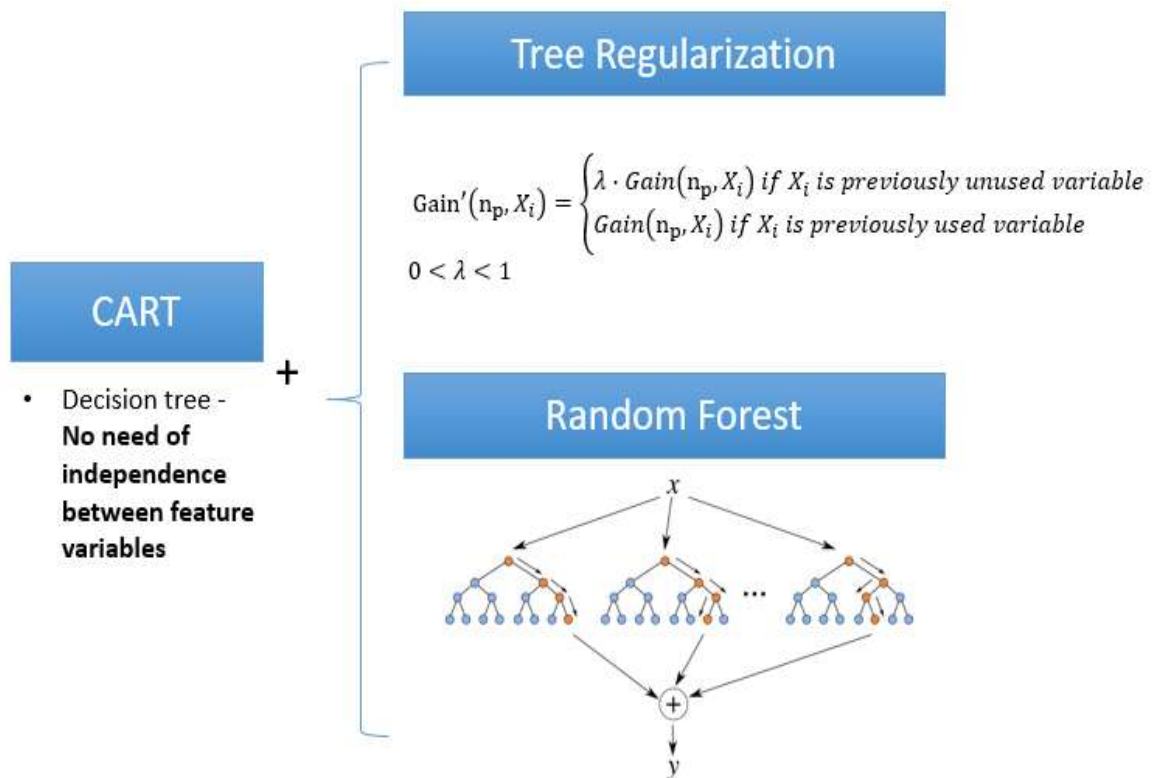


그림 31 데이터 마이닝 기법 적용 방법,

출처: Random Forest, <http://kazoo04.hatenablog.com/entry/2013/12/04/175402>

- 일반적으로 변수들이 종속변수에 미치는 영향을 측정하기 위해 통계적 분류 모델이 사용됨
- 사고 심각도에 대한 risk factor의 영향을 확인하고 분석하기 위하여 본 연구에서는 주어진 feature variable (risk factor)에 대한 class variable (collision severity)을 예측하는 통계적 분류 모델을 수립함
- 선행 연구들에 의하면 주요 사고 위험 요인간에는 상호 연관성이 존재하며, 이를 고려한 데이터마이닝 분류 기법인 classification and regression tree, 즉 CART를 기반으로 Tree Regularization과 Random Forest 기법을 활용함
- Tree Regularization 기법은 한번 사용된 split variable을 최대한 사용하고, 최소한의 split variables를 사용하게 하는 기법으로 이를 통해 risk factor의 상대적 중요도 차이를 더욱 분명하게 할 수 있음
- Random Forest는 모델의 전체적 정확도를 향상시키는 역할을 함

(1) Regularized Decision Tree

- CART는 classification 문제에 있어서 가장 널리 쓰이는 데이터마이닝 기법 중 하나임

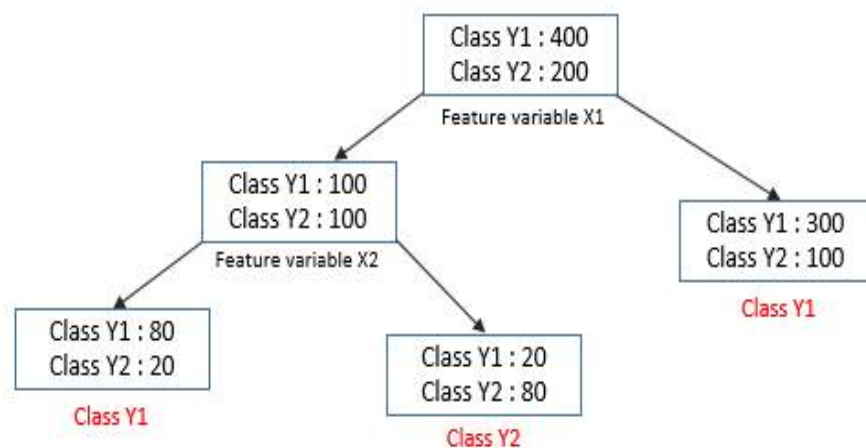


그림 32 Classification 예시

- CART를 통해 아래의 수식과 같이 지니 불순도 (Gini Impurity)를 계산하고 split variable을 선택하게 됨
- Gini impurity는 불순도를 측정하는 지수로서 Gini 지수를 감소시키는 예측변수와 그때의 최적분리에 의해서 child node를 선택하며 계산식은 다음과 같음

$$Gini(n) = \sum_{i=1}^m Y_i(1 - Y_i)$$

n: node, m: number of class variable, Y_i : proportion of i th class variable in the node

- 본 연구에서 사용된 m 값은 fatal, injury, PDO 3가지임
- 만약 parent node n_p 가 feature variable X에 의해 분리될 때, Gini index의 감소값이 Gain 값이 되며 다음과 같이 계산됨

$$Gain(n_p, X) = Decreased\ Gini(X) = Gini(n_p) - \sum_{i=1}^k \alpha_i Gini(n_{c,i})$$

n_p : split by a feature variable X

k: number of child nodes

$n_{c,i}$: i th child node split by variable X

α_i : proportion of the number of data tuples in i th child node to the number of data tuples in parent node

- parent node n_p 가 주어지고 feature variable $X = \{X_1, X_2, \dots, X_j, \dots\}$ 일 때, information gain의 변형 모델은 아래와 같이 계산되며 λ 는 splitting variable로 사용되지 않은 feature variable에 대한 penalty로서 0부터 1 사이의 값을 사용함

$$Gain'(n_p, X_i) = \begin{cases} \lambda \cdot Gain(n_p, X_i) & \text{if } X_i \text{ is previously unused variable} \\ Gain(n_p, X_i) & \text{if } X_i \text{ is previously used variable} \end{cases}$$

- Regularized Tree 기법을 통해 아래 예시와 같이 더 정확한 feature variable을 도출할 수 있으나 그로 인해 전체적인 예측 성능은 저하될 수 있음

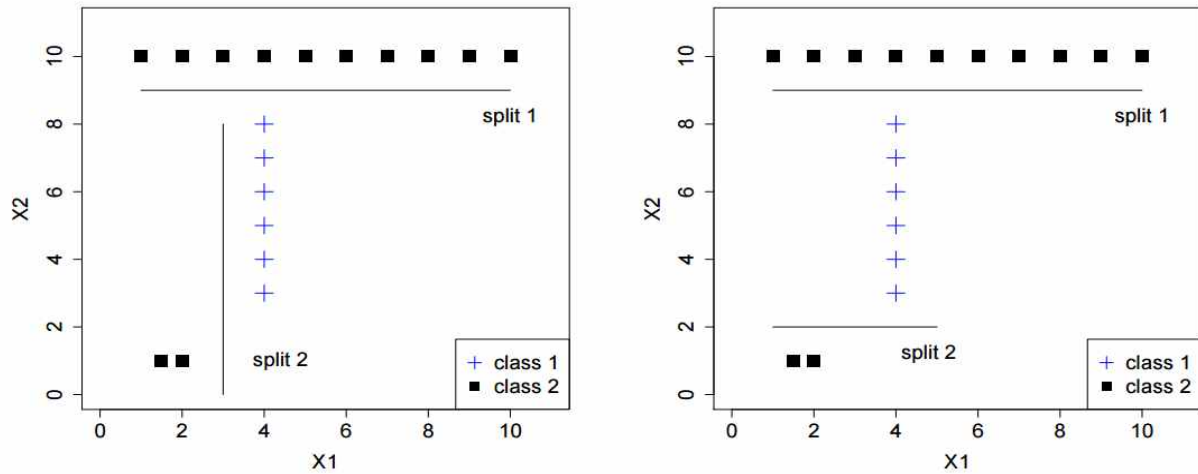


그림 33 Regularized tree를 통한 feature selection 변화 예시

출처; Deng, H., & Runger, G. (2012, June). Feature selection via regularized trees. In *Neural Networks (IJCNN), The 2012 International Joint Conference on* (pp. 1-8). IEEE.

(2) Random Forest

- Random Forest 모형은 Decision Tree의 확장 모델로서 동일한 하나의 데이터 집합에서 임의의 복원 샘플링을 통해 여러 개의 훈련용 데이터를 만들어 여러 번의 학습을 통해 트리를 생성하고 이들을 결합해 최종적으로 목표 변수를 예측함
- 이 모형은 앙상블 기법인 배깅 (bagging) 방법과 임의의 노드 최적화를 통해 랜덤하게 분석에 사용되는 변수를 추출할 수 있음
- Decision Tree는 노드를 결정짓게 되는 매개변수의 영향이 크므로 깊이가 깊은 단일 Decision Tree의 경우 과대적합 (Overfitting)의 위험이 큼
- 그러므로 배깅을 이용해 각 트리의 확률을 통해 목표변수를 예측하는 Random forest의 경우, 트리들의 편향을 그대로 유지하되 분산은 감소하므로 안정적이며 정확도가 높아지게 됨

(3) 분석 툴

- 모델 학습을 위해 프로그래밍 툴인 R을 사용하였으며 RRF 패키지를 적용하였음
- 각 위험 요인의 중요도는 mean decrease accuracy를 기준으로 순위를 매김
- Mean decrease accuracy란 각 트리에서 주어진 변수가 사용될 때와 그렇지 않을 때의 모델 정확도 차이를 Random Forest 내에 존재하는 트리의 수로 평균을 내어 계산함
- 이러한 작업을 통해, 결과적으로 사고 위험 요인 중요도에 대한 순위표를 도출함

(4) Dissimilarity Index

- Subgroup의 사고 risk factor 순위 차이를 정량적으로 비교하기 위해 dissimilarity index를 고안함
- Dissimilarity index는 하나의 subgroup의 risk factor별 순위를 순서대로 나열한 것을 하나의 벡터로 취급하고, 두 subgroup의 risk factor 순위 벡터의 유클리드 거리를 구하여 계산함
- 두 subgroup의 risk factor 순위표가 일치하는 경우 최소값 0을 나타내며, 차이가 클수록 큰 값을 가지게 됨
- 두 subgroup간의 dissimilarity index가 높을수록 충돌 상황에서 두 집단의 행동 양상과 심각도 결정 규칙이 상이함을 의미함
- Dissimilarity index는 아래의 수식을 통해 계산할 수 있음

Rank vector of subgroup $X \equiv (R_{rf1}, R_{rf2}, \dots, R_{rfn})$

rf_k : k_{th} risk factor, n : number of risk factors

$R_{rfk} \in (1, 2, 3, \dots, n)$: Importance Rank of rf_k

Dissimilarity index of two Rank vector

$$DI(X, Y) = \sqrt{\sum_{i=1}^n (X(R_{rfi}) - Y(R_{rfi}))^2}$$

- 예를 들어, 남성 운전자와 여성 운전자의 Dissimilarity를 구하기 위해서는 아래의 표와 같이 사고 위험 요인에 대한 순위가 형성됨
- 남성 운전자의 경우 1, 2, 3, 4, ..., 10의 벡터가 형성되고 여성 운전자의 경우도 2, 1, 4, 3, ..., 9의 벡터가 형성됨
- 이 두 벡터의 유클리드 거리를 구하면 비유사도 값인 2.83이 도출됨

Rank vector of Male = $X^{\text{Male}} = (R_{\text{VIOLCAT}}, R_{\text{CRASHTYP}}, \dots, R_{\text{VEHAGE_Group}}) = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$

Rank vector of Female = $X^{\text{Female}} = (R_{\text{VIOLCAT}}, R_{\text{CRASHTYP}}, \dots, R_{\text{VEHAGE_Group}}) = (1, 5, 2, 4, 3, 6, 7, 8, 9, 10)$

$$\text{Dissimilarity}(X^{\text{Male}}, X^{\text{Female}}) = [(1-1)^2 + (2-5)^2 + \dots + (10-10)^2]^{1/2} = 2.83$$

Risk Factor	Rank	
	Male	Female
VIOLCAT	1	2
CRASHTYP	2	1
MOVEMENT	3	4
STATEHW	4	3
POP	5	5
ATFAULT	6	6
PSOBER	7	8
PRACE	8	7
PAGE_Group	9	10
VEHAGE_Group	10	9

그림 35 Dissimilarity Index를 통한 순위표 예시

사. 연구 결과

(1) 연구 결과 요약

- 앞서 언급한 각 subgroup에 대하여, 다음의 표와 같은 사고 risk factor 순위표가 도출됨

Rank	Gender of driver		Population		At-fault party indicator	
	Male	Female	Urban	Rural	At-fault	Not-at-fault
1	VIOLCAT	CRASHTYP	CRASHTYP	VIOLCAT	CRASHTYP	CRASHTYP
2	CRASHTYP	VIOLCAT	STATEHW	CRASHTYP	VIOLCAT	VIOLCAT
3	MOVEMENT	STATEHW	VIOLCAT	MOVEMENT	MOVEMENT	STATEHW
4	STATEHW	MOVEMENT	MOVEMENT	ATFAULT	PSOBER	POP
5	POP	POP	ATFAULT	PSOBER	POP	MOVEMENT
6	ATFAULT	ATFAULT	PSEX	STATEHW	STATEHW	PRACE
7	PSOBER	PRACE	PSOBER	PSEX	PRACE	VEHAGE_Group
8	PRACE	PSOBER	PRACE	PAGE_Group	PSEX	PAGE_Group
9	PAGE_Group	VEHAGE_Group	VEHAGE_Group	VEHAGE_Group	PAGE_Group	PSEX
10	VEHAGE_Group	PAGE_Group	PAGE_Group	PRACE	VEHAGE_Group	PSOBER

Rank	Race of Driver				Vehicle Maker			
	White	Black	Hispanic	Asian	Toyota	Ford	BMW	Hyundai
1	CRASHTYP	CRASHTYP	VIOLCAT	CRASHTYP	CRASHTYP	CRASHTYP	VIOLCAT	VIOLCAT
2	VIOLCAT	VIOLCAT	CRASHTYP	VIOLCAT	VIOLCAT	VIOLCAT	STATEHW	CRASHTYP
3	MOVEMENT	MOVEMENT	MOVEMENT	MOVEMENT	STATEHW	MOVEMENT	CRASHTYP	STATEHW
4	STATEHW	STATEHW	POP	STATEHW	MOVEMENT	STATEHW	POP	MOVEMENT
5	POP	ATFAULT	STATEHW	POP	POP	POP	MOVEMENT	POP
6	ATFAULT	POP	ATFAULT	ATFAULT	ATFAULT	ATFAULT	ATFAULT	ATFAULT
7	PSOBER	PSOBER	PSOBER	PSOBER	PRACE	PSOBER	PRACE	PRACE
8	PAGE_Group	PSEX	PSEX	PAGE_Group	VEHAGE_Group	PRACE	PSOBER	PAGE_Group
9	VEHAGE_Group	VEHAGE_Group	VEHAGE_Group	VEHAGE_Group	PSOBER	PSEX	VEHAGE_Group	PSEX
10	PSEX	PAGE_Group	PAGE_Group	PSEX	PAGE_Group	VEHAGE_Group	PSEX	PSOBER
11					PSEX	PAGE_Group	PAGE_Group	VEHAGE_Group

Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 36. Risk factor 순위표 요약

- 분류 기준이 되는 14개의 각 subgroup에 대해 살펴보면, 운전자 성별 (Male, Female)에 의해 만들어진 두 subgroup을 보면, 전체적인 순위 경향성이 매우 유사함을 알 수 있음
- 또한, 주요 위험 요인이라 할 수 있는 5위 이내에 위치한 위험 요인들이 두 집단에서 동일함
- 특히 VIOLCAT, CRASHTYP, MOVEMENT, STATEHW, POP은 아래와 같이 전체 데이터에 대해서도 가장 중요한 risk factor로 도출됨

Risk Factor	Mean Rank	Standard Deviation	Rank in Whole Data
CRASHTYP	1.6364	0.6544	1
VIOLCAT	1.8864	0.6558	2
MOVEMENT	3.5114	1.0533	4
STATEHW	3.7614	1.5636	3
POP	4.5845	1.2874	5
ATFAULT	5.7500	0.4330	6
PRACE	6.7653	2.2545	7
PSOBER	6.8239	2.3089	9
PSEX	7.9416	2.5739	10
VEHAGE_Group	8.3864	2.3459	8
PAGE_Group	8.5114	2.4500	11

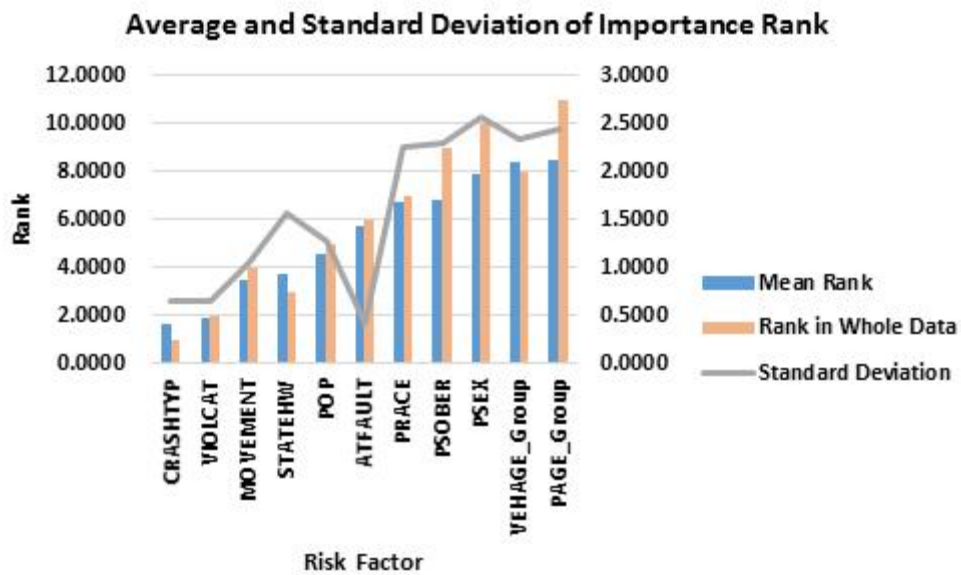
Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 37. Risk Factor 순위표의 기초 통계 분석 결과

Rank	Entire population
1	CRASHTYP
2	VIOLCAT
3	STATEHW
4	MOVEMENT
5	POP
6	ATFAULT
7	PRACE
8	VEHAGE_Group
9	PSOBER
10	PSEX
11	PAGE_Group

Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 38. 전체 Population 에 대한 순위표

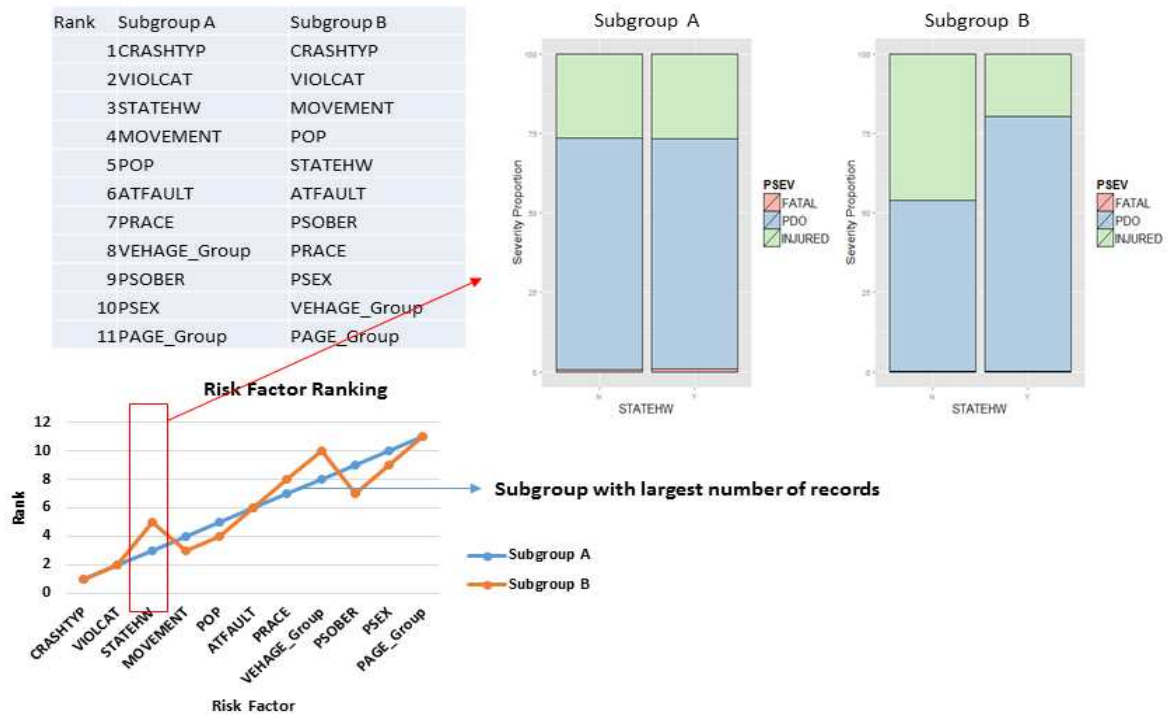


Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 39. Risk factor 순위표에 대한 평균 및 표준편차

-
- 다른 subgroup 기준에 대해서는 top 5 risk factor가 전체 데이터의 경우와 일치하지 않는 경우도 존재함
 - 도시와 시골지역 (Highly urban, Rural)의 risk factor 순위 차이를 보면, 주요 risk factor들 사이에 차이가 있음을 알 수 있음
 - 먼저 STATEHW가 도심에서는 두 번째로 중요한 반면, 시골에서는 6위로 순위가 하락함
 - VIOLCAT의 경우 양쪽 모두 주요 위험 요인에 해당되지만, 일반적으로 매우 중요한 risk factor인 만큼 1위와 3위의 차이도 상대적으로 주목할 만 하다고 판단됨
 - 피해차량과 가해차량(At-fault, Not-at-fault)의 경우 1위와 2위의 순위는 일치하나 3위부터 상당히 다른 양상을 보임
 - MOVEMENT는 양쪽 모두에서 5위 안에 드는 중요 risk factor라고 할 수 있으나 가해차량에서 2순위만큼 그 중요도가 높아짐
 - 또한 SOBER의 경우에는 가해차량에서는 4위지만, 피해차량에서는 가장 영향력이 없는 risk factor로 도출됨
 - 반대로 피해 차량에서 중요한 risk factor로 도출된 STATEHW는 가해차량에서는 순위가 6위임을 확인함
 - 인종(White, Black, Hispanic, Asian)별 차이를 살펴보면, 백인 subgroup과 아시안 subgroup의 순위가 일치함
 - 또한 5위안에 들어가는 주요 risk factor들만 놓고 보았을 때, 백인, 히스패닉, 아시안의 경우 같은 risk factor들을 공유하였지만 흑인의 경우 POP 대신 ATFAULT가 위치함

(2) 결과 분석 기준



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 40. 분석 결과 해석 방법

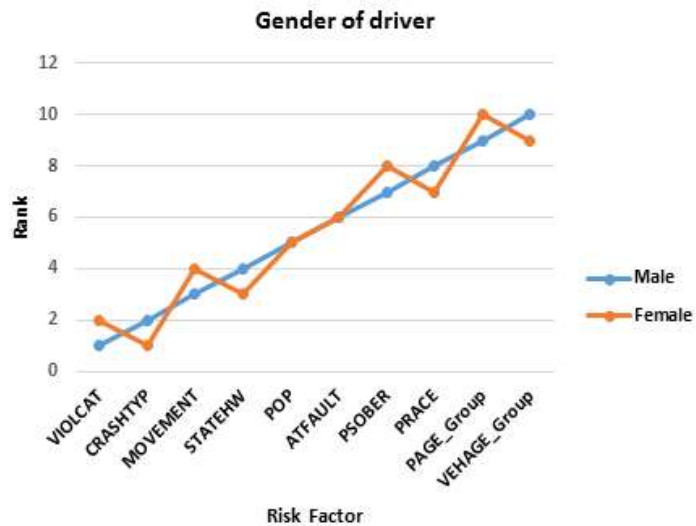
- 결과에 대한 설명에 앞서 subgroup 분류 기준 별 상세 분석 과정에서의 결과 분석 기준은 다음과 같음
- Subgroup 분류 기준을 통해 생성되는 risk factor 순위는 표를 통해 나타냄
- 또한, 순위 경향성 파악을 위해 risk factor 순위를 그래프로 나타내며 이때 기준선, 즉 x 축의 risk factor 나열 기준은 subgroup 중 가장 많은 기록을 가지고 있는 subgroup 을 기준으로 결정됨
- risk factor 중 두 subgroup에서의 순위 차이가 2 이상이며, 하나 이상의 subgroup에서 순위가 6위보다 같거나 높은 경우에는 오른쪽 그림과 같이 실제 심각도 분포를 도출함
- 순위 차이가 2 이상이 나는 경우라도, 모든 subgroup에서 순위가 6위보다 낮은 경우에는 심각도 분포를 측정하지 않으며, 이는 risk factor의 중요도가 6위 이후로 급격하게 감소하기 때문에 실질적인 영향력이 미미하다고 판단하였기 때문임
- 또한, 예외적으로 VIOLCAT, 즉 법규 위반의 경우 25개로 너무 많은 항목들이 존재하여 순위차이가 존재하더라도 심각도 분포를 측정하지 않음
- Subgroup이 4개 존재하는 운전자 인종 별 subgroup이나 차량 제조사별 subgroup의 경우, 평균 순위 차이가 2 이상이고 순위가 6위 이상인 사고 위험 요인에 대해 실제 분포를 측정하려 하였으나 해당 조건을 만족하는 위험 요인이 존재하지 않음

(3) Subgroup 별 risk factor 순위 결과

(가) Gender of driver

Rank Male	Female
1 VIOLCAT	CRASHTYP
2 CRASHTYP	VIOLCAT
3 MOVEMENT	STATEHW
4 STATEHW	MOVEMENT
5 POP	POP
6 ATFAULT	ATFAULT
7 PSOBER	PRACE
8 PRACE	PSOBER
9 PAGE_Group	VEHAGE_Group
10 VEHAGE_Group	PAGE_Group

$$DI(\text{Male}, \text{Female}) = 2.83$$



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 41. Subgroup 별 분석 결과: 운전자 성별

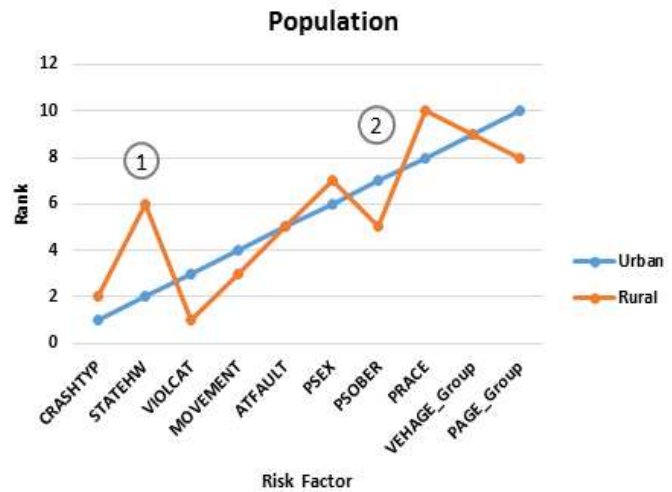
- 운전자 성별에 따른 risk factor의 순위 분포는 Dissimilarity index 2.83으로 거의 일치함
- 1위와 2위, 3위와 4위, 7위와 8위, 9위와 10위에서 각 subgroup의 risk factor 순위 변동이 존재하나 2이상의 눈에 띄는 순위 차이를 보이는 위험 요인은 존재하지 않음

(나) Population

- Population에 따른 도시와 시골의 subgroup에 대한 dissimilarity index는 6.63으로 비교적 큰 차이를 보임
- 특히, STATEHW 즉 주 고속도로 여부가 도시에서는 두 번째로 중요한 risk factor인 반면 시골에서는 6번째로 중요한 risk factor로 나타남
- 그 외, 운전자 성별 (PSEX)와 인종 (PRACE)이 도시에서 시골보다 중요한 risk factor로 나타난 반면, 시골에서는 VIOLCAT, MOVEMENT, PSOBER, PAGE_Group이 도시보다 중요한 risk factor로 나타남
- 특히 ①STATEHW와 ②PSOBER는 주목할 만한 순위 변화를 나타냈으며 해당 risk factor에 대한 심각도 분포를 살펴봄

Rank Urban	Rural
1 CRASHTYP	VIOLCAT
2 STATEHW	CRASHTYP
3 VIOLCAT	MOVEMENT
4 MOVEMENT	PSOBER
5 ATFAULT	ATFAULT
6 PSEX	STATEHW
7 PSOBER	PSEX
8 PRACE	PAGE_Group
9 VEHAGE_Group	VEHAGE_Group
10 PAGE_Group	PRACE

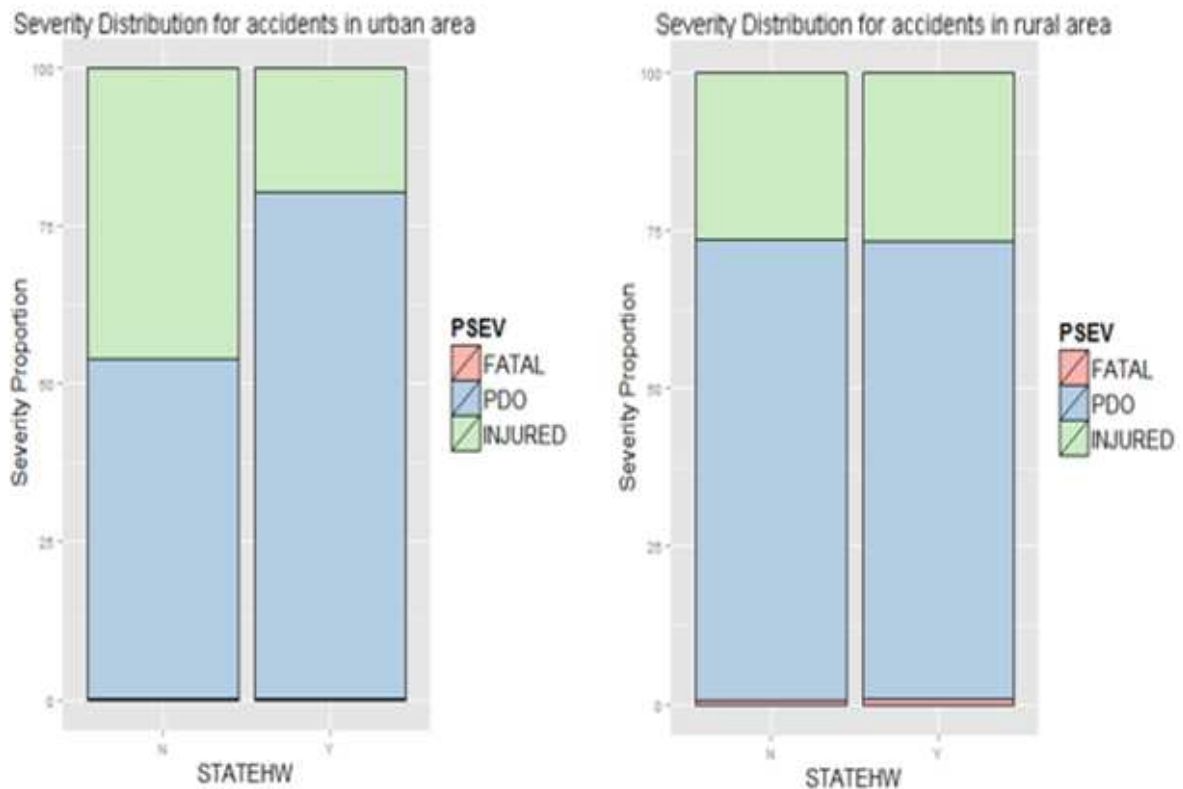
$$DI(Urban, Rural) = 6.63$$



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 42. Subgroup 별 분석 결과: 인구

① STATEHW의 심각도 분포



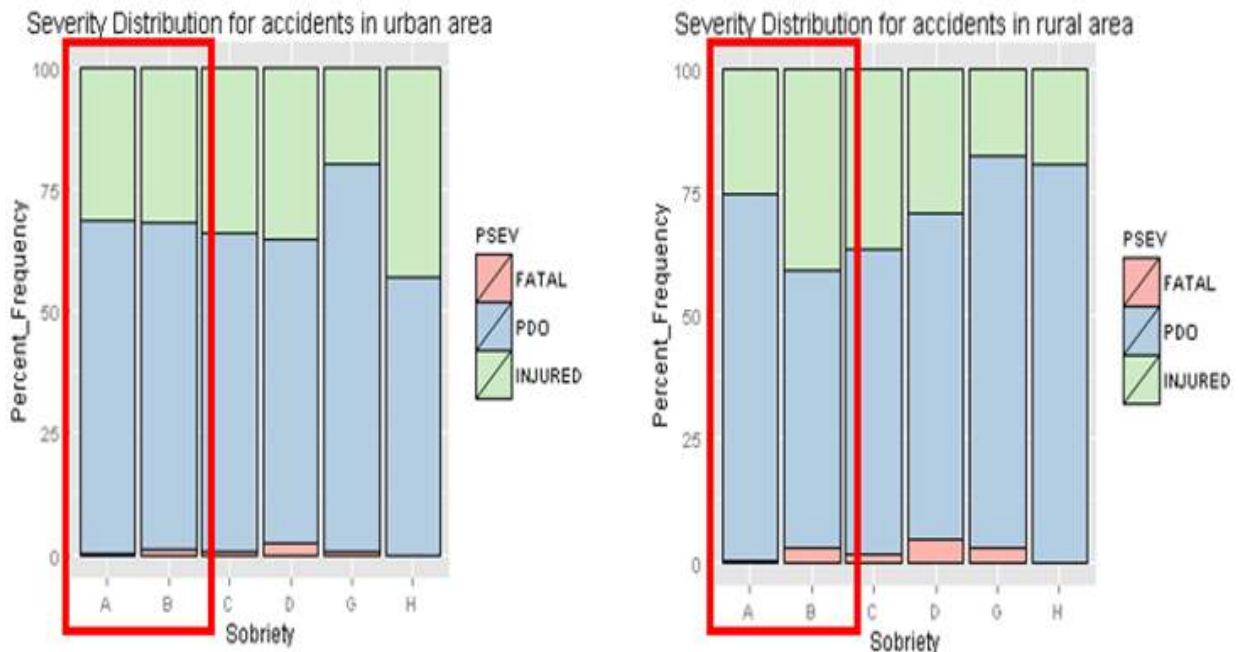
Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 43. 인구차이 별 주 고속도로에 따른 심각도 분포

- 두 그래프에서 주 고속도로 여부에 의해 충돌 시 사상 발생 확률의 변화 정도가 현격히 차이가 있음을 알 수 있음

- 도시 지역에서는 차량 충돌 시 사상 확률, 즉 사망하거나 부상사고가 일어날 확률이 주 고속도로에서는 19.81%에 그치는 한편 주 고속도로가 아닌 곳에서는 46.25%로 매우 증가함
- 반면 지방 혹은 시골 지역에서는 차량 충돌 시 사상 확률이 주 고속도로가 아닌 곳과 주 고속도로인 곳에서 각각 26.96%와 27.34%로 거의 동일함
- 각 지역에서의 주 고속도로 및 주 고속도로 아닌 곳에서의 통행량이나 교통 흐름의 차이 정도에 기인하는 것으로 판단됨

② PSOBER의 심각도 분포



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 44. 인구차이 별 음주운전에 따른 심각도 분포

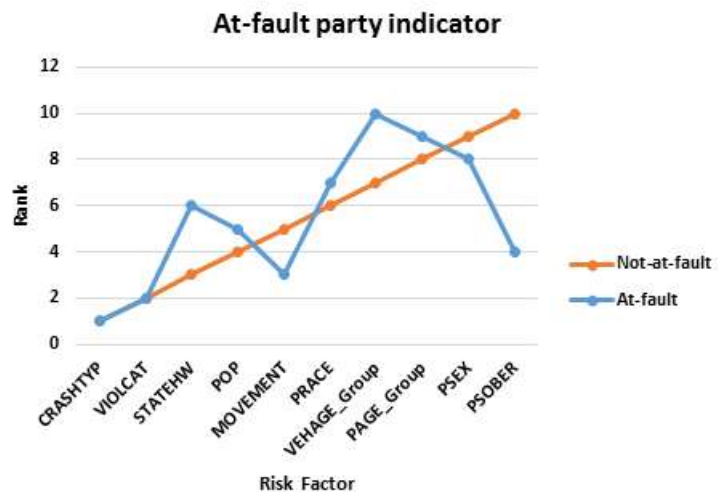
- 음주여부에 따른 심각도 분포를 살펴보기에 앞서 음주운전에 대한 구분은 다음과 같음
- A: not under influence of alcohol
- B: under influence of alcohol
- C, D, G, H: minor categories such as cases where driver was drunk driving but not under influence of alcohol
- 음주 운전 여부 항목에 있어 사고 기록의 대부분은 차지하는 것은 A와 B이며, 이는 각각 취하지 않은 운전자와 음주 운전자를 의미함

- 도시 지역에서는 비 음주 운전자의 사상 확률이 31.43%에 비해 음주 운전자의 사상 확률은 33.14%로 소폭 증가함
- 반면, 시골 지역에서는 비 음주 운전자의 사상 확률이 25.82%로 도시보다도 낮은 반면 음주 운전시 충돌 후 사상 확률이 43.90%로 매우 증가함
- 도시지역의 높은 교통량이 음주 운전의 충돌 시 사상 확률 증가 폭을 경감시키는 것으로 판단됨

(다) At-fault party indicator

Rank Not-at-fault	At-fault
1 CRASHTYP	CRASHTYP
2 VIOLCAT	VIOLCAT
3 STATEHW	MOVEMENT
4 POP	PSOBER
5 MOVEMENT	POP
6 PRACE	STATEHW
7 VEHAGE_Group	PRACE
8 PAGE_Group	PSEX
9 PSEX	PAGE_Group
10 PSOBER	VEHAGE_Group

$$DI(Not_at_fault, At_fault) = 7.87$$

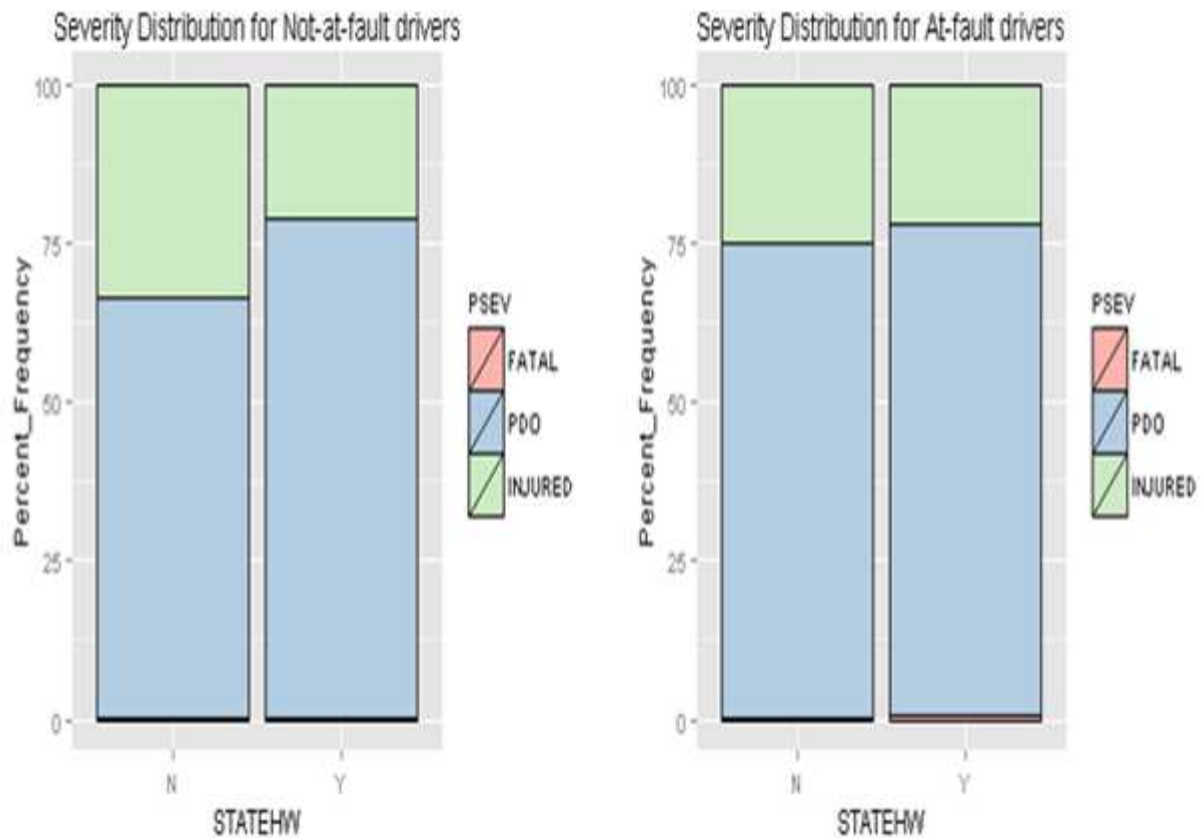


Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 45.Subgroup 별 분석 결과: 가해차량 여부

- 가해 차량과 피해 차량의 risk factor 순위 분포 차이는 subgroup 분류 기준 중 가장 큰 값을 보임
- 그래프에서 그 차이를 확인할 수 있으며 피해 차량에 두드러진 영향력을 미치는 risk factor로는 STATEHW를 꼽을 수 있으며, POP, PRACE, VEHAGE_Group, PAGE_Group 역시 가해 차량에서보다 높은 중요도를 보임
- 반면, 가해 차량에서 두드러지는 영향력을 행사하는 위험 요인으로는 MOVEMENT와 PSOBER가 있으며 특히 PSOBER는 6의 순위 차이를 보임
- PSEX 역시 낮은 순위지만 순위 차이를 보이고 있음

① STATEHW의 심각도 분포

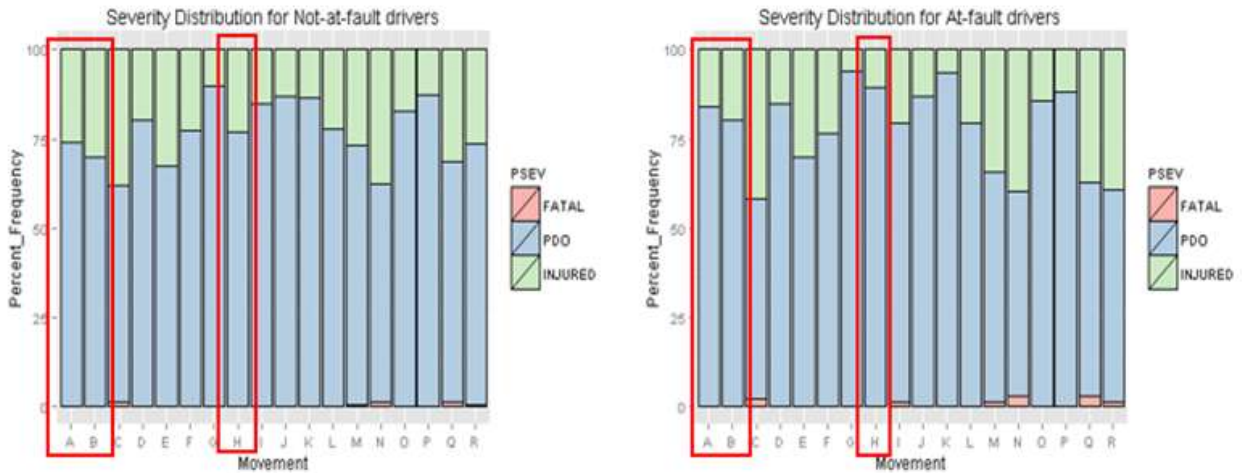


Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 46. 가해차량 여부 별 주 고속도로에 따른 심각도 분포

- 앞서 언급된 인구수에 따른 변화와 동일하게 STATEHW가 아닌 곳에서 항상 심각한 사고가 더 빈번히 발생함
- 피해차량의 경우 고속도로인 경우와 아닌 경우의 심각도 분포 차이가 각각 21.27%, 33.56%로 가해 차량의 경우인 22.49%와 25.49%보다 더 크게 나타남

② MOVEMENT의 심각도 분포

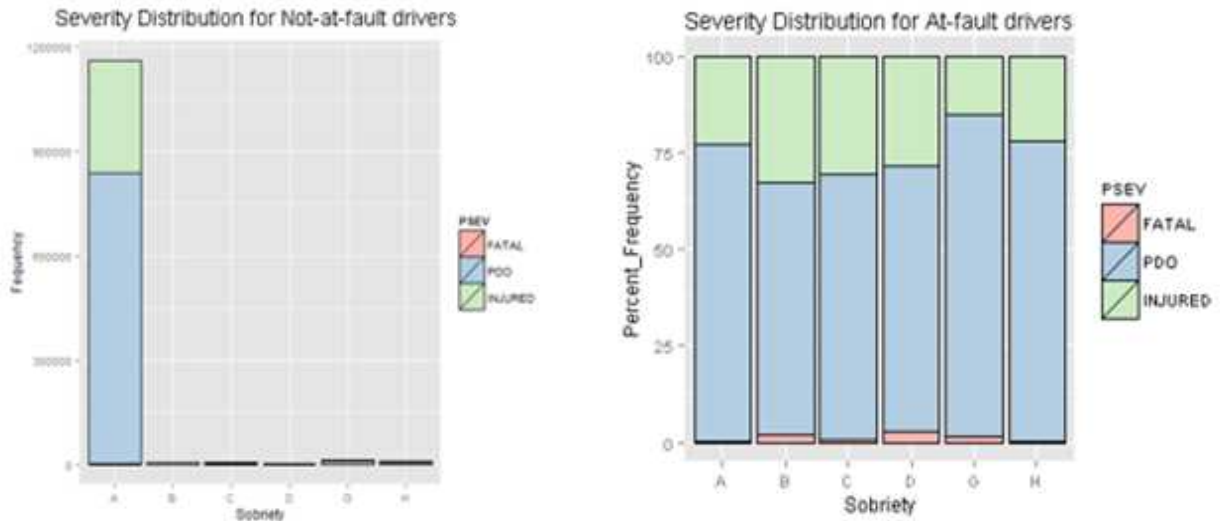


Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 47. 가해차량 여부 별 Movement에 따른 심각도 분포

- MOVEMENT의 경우 피해 차량에서 A, B, H 항목에 대해서 가해차량에 비해 낮은 분포를 보임 (A - Stopped, B - Proceeding Straight, H - Slowing/Stopping)

③ PSOBBER의 심각도 분포



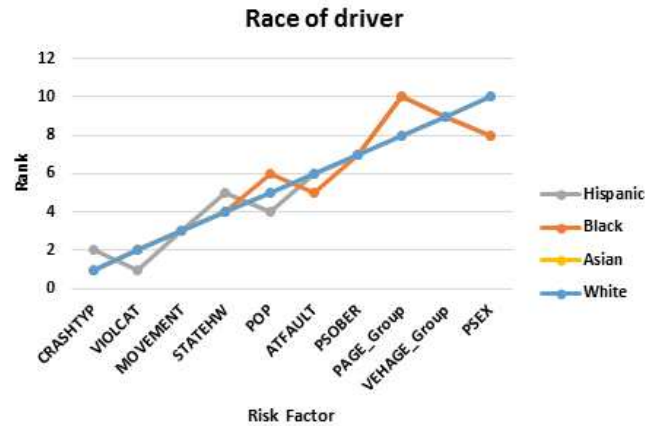
Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 48. 가해차량 여부 별 음주 운전 여부에 따른 심각도 분포

- 음주운전 여부와 관련하여, 사고에 있어서 피해 차량이 음주 운전의 영향 하에 놓일 확률은 매우 적음
- 실제 분포를 살펴보면, 98.1%로 대다수의 피해 차량은 음주 운전의 영향이 없음을 확인할 수 있음
- 따라서 음주운전 여부가 주요한 risk factor로서 작용할 수 없음

(라) Race of driver

Rank	White	Black	Hispanic	Asian
1	CRASHTYP	CRASHTYP	VIOLCAT	CRASHTYP
2	VIOLCAT	VIOLCAT	CRASHTYP	VIOLCAT
3	MOVEMENT	MOVEMENT	MOVEMENT	MOVEMENT
4	STATEHW	STATEHW	POP	STATEHW
5	POP	ATFAULT	STATEHW	POP
6	ATFAULT	POP	ATFAULT	ATFAULT
7	PSOBER	PSOBER	PSOBER	PSOBER
8	PAGE_Group	PSEX	PSEX	PAGE_Group
9	VEHAGE_Group	VEHAGE_Group	VEHAGE_Group	VEHAGE_Group
10	PSEX	PAGE_Group	PAGE_Group	PSEX



$DI(White, Black) = 3.16$
 $Max\ DI(White, Hispanic) = 3.46$
 $Min\ DI(White, Asian) = 0$
 $DI(Black, Hispanic) = 2.83$
 $DI(Black, Asian) = 3.16$
 $Max\ DI(Hispanic, Asian) = 3.46$

Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

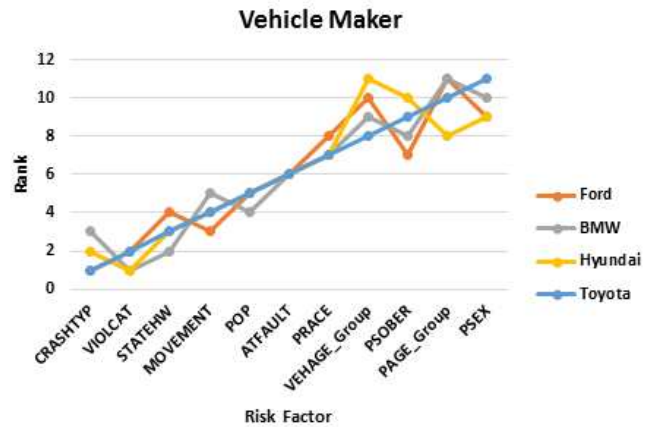
그림 49. Subgroup 별 분석 결과: 인종

- 인종 별 subgroup에 대한 risk factor 순위 변화를 살펴보면, 4개의 subgroup에 대해 6개의 dissimilarity index가 계산되었고 그 결과, 가장 유사한 subgroup은 DI값이 0으로 백인과 아시아인이었으며 비 유사도가 높은 두 subgroup은 3.46으로 히스패닉과 백인 혹은 히스패닉과 아시아인으로 나타남

(마) Vehicle maker

Rank	Toyota	Ford	BMW	Hyundai
1	CRASHTYP	CRASHTYP	VIOLCAT	VIOLCAT
2	VIOLCAT	VIOLCAT	STATEHW	CRASHTYP
3	STATEHW	MOVEMENT	CRASHTYP	STATEHW
4	MOVEMENT	STATEHW	POP	MOVEMENT
5	POP	POP	MOVEMENT	POP
6	ATFAULT	ATFAULT	ATFAULT	ATFAULT
7	PRACE	PSOBER	PRACE	PRACE
8	VEHAGE_Group	PRACE	PSOBER	PAGE_Group
9	PSOBER	PSEX	VEHAGE_Group	PSEX
10	PAGE_Group	VEHAGE_Group	PSEX	PSOBER
11	PSEX	PAGE_Group	PAGE_Group	VEHAGE_Group

$DI(\text{Toyota}, \text{Ford}) = 4.00$
Min $DI(\text{Toyota}, \text{BMW}) = 3.46$
 $DI(\text{Toyota}, \text{Hyundai}) = 4.47$
 $DI(\text{Ford}, \text{BMW}) = 4.24$
Max $DI(\text{Ford}, \text{Hyundai}) = 4.90$
 $DI(\text{BMW}, \text{Hyundai}) = 4.69$



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 50. Subgroup 별 분석 결과: 차량 제조사

- 차량 제조사에 의한 평균 dissimilarity는 운전자 정보에 의한 차이보다 다소 높은 값을 형성하였으며, 이에 따라 순위 변화 그래프도 전체적으로는 유사하나 성별이나 인종에 따른 변화보다는 좀 더 큰 변화폭을 보이고 있음을 확인할 수 있음
- 제조사 간 risk factor 순위가 가장 비슷한 두 subgroup은 토요타와 BMW였으며, 가장 큰 두 그룹은 현대와 포드로 나타남

(4) 결과 요약

Subgrouping Criteria	X	Y	$DI(X, Y)$
Gender of driver	Male	Female	2.83
Population	Urban	Rural	6.63
At-fault party indicator	Not-at-fault	At-fault	7.87

Matrix of $DI(X, Y)$						
Subgroup by Race of driver		Y				Average
		White	Asian	Black	Hispanic	
X	White		0	3.16	3.46	2.21
	Asian		0	3.16	3.46	2.21
	Black	3.16	3.16		2.83	3.05
	Hispanic	3.46	3.46	2.83		3.25
Subgroup by vehicle maker		Y				Average
		Toyota	Ford	BMW	Hyundai	
X	Toyota		4.00	3.46	4.47	3.98
	Ford	4.00		4.24	4.90	4.38
	BMW	3.46	4.24		4.69	4.13
	Hyundai	4.47	4.90	4.69		4.69

Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 51. Dissimilarity Index를 통한 비교 분석 결과 요약

- Subgroup간 차이를 정량적으로 계산하기 위해 정의했던 dissimilarity index에 대한 결과 요약은 다음과 같음
- 운전자 인종, 차량 제조사의 경우 4개의 subgroup이 생성되며, dissimilarity는 두 집단 간에 계산되므로 다음과 같은 matrix 형태로 나타나게 됨
- 이 경우에는 마지막 열에 평균값을 표기하였으며 이 값은 하나의 subgroup이 나머지 subgroup들이 가지는 평균 비유사도로 판단할 수 있음
- 따라서 높은 평균 dissimilarity를 가지는 subgroup은 전체 그룹 내에서 가장 이질적인 사고 심각도 결정 규칙을 가짐을 판단할 수 있음

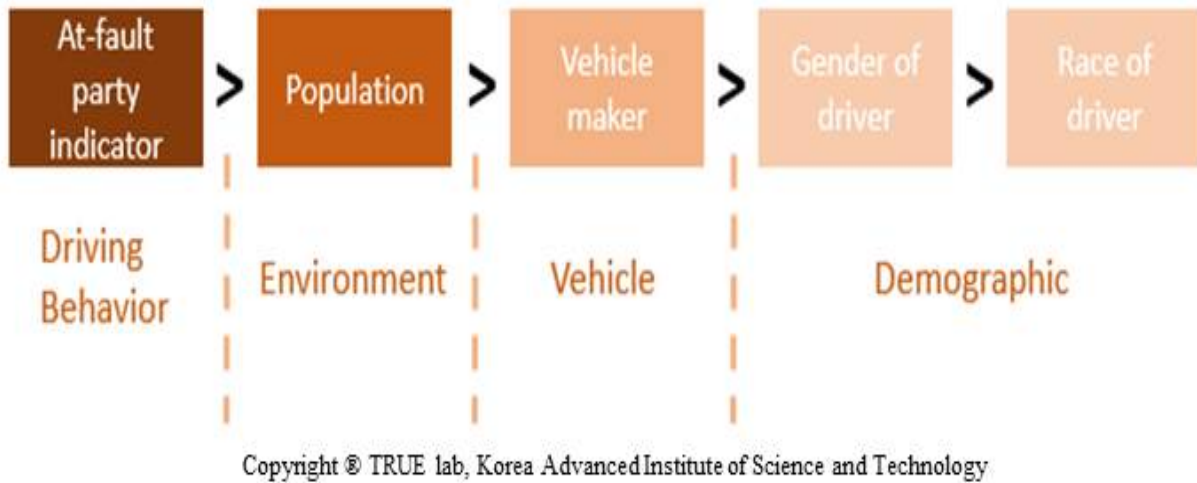


그림 52. Subgroup 분류 기준에 따른 평균 비유사도 비교 결과

- 종합적으로, 각 subgroup 분류 기준별로 생성되는 subgroup dissimilarity index의 평균치를 보았을 때, 다음과 같은 차이를 나타냄
- 가해 차량 여부로 분류했을 때 가장 큰 사고 risk factor 중요도 분포 차이를 보이며, 그 다음으로 도시와 시골의 차이가 컸고, 차량 제조사에 의한 차이, 운전자 성별에 의한 차이, 운전자 인종에 의한 차이 순이었음
- 앞서 분류 기준 선정과정에서 언급했듯, 다섯 가지 분류 기준들은 운전자 변수, 차량 변수, 외부 환경적 변수, 행동 변수로 나눌 수 있음
- 본 연구 결과는 운전자나 차량에 의한 차이보다는 외부 환경적 요인이나 사고와 직접적으로 관련된 행동 양식의 차이가 심각도 결정 양상에 더 큰 차이를 보임을 정량적으로 증명할 수 있음을 나타냄

아. 결론

- 본 연구는 대용량 사고 데이터로부터 다양한 하위그룹을 생성한 후 각각의 risk factor를 찾아내고, 그룹별로 비교하여 그 차이를 측정하는 것을 목적으로 함
- 연구를 위해 SWITRS의 데이터가 사용되었으며, 운전자 성별, 인구, 가해차량여부, 인종, 차량 제조사 별로 subgroup을 생성함
- 각 subgroup에 regularized random forest 기법을 적용하여 주요 위험 요인을 찾아내고, 중요도 순위를 측정하여 비교 분석함
- Subgroup 간의 정량적 차이를 측정하기 위해 dissimilarity index를 고안하여 도입하였으며, 그 결과 주요 risk factor 차이가 운전 행동양식 및 외부 환경적 차이에 의해 많은 영향을 받으며 운전자나 차량 정보에 의해서는 다소 차이를 보이지 않음을 확인함
- 앞선 문헌조사에서 본 연구의 차별점을 언급한 바와 같이, 다양한 특징을 가진 집단들 간의 risk factor 차이를 분석한 것과 risk factor간 상호의존성을 고려한 분석 방법론 두 가지 조건을 동시에 만족하는 연구임
- 또한 subgroup 간의 risk factor 중요도 차이를 계산하는 dissimilarity index 및 정성적 분석 뿐 아니라 정량적 분석을 수행하였으며 본 방법론은 기타 다른 데이터에서 적용될 수 있음

제4장. 목표 달성도 및 관련 분야 기여도

1. 목표 달성도

연구 내용				달성도 (%)
A-1	교통사고 위험요인 분석을 위한 관련 데이터 통합 기술	A-1-①	활용 가능 데이터베이스 조사	100 %
		A-1-②	통합 데이터 플랫폼 개발	100 %
		A-1-③	데이터 통합기술 검증	100 %
A-2	교통사고 심각도 결정 모델	A-2-①	사고 심각도 분석 선행 연구 조사	100 %
		A-2-②	위험요인 분석을 위한 방향성 제시	100 %
		A-2-③	데이터 마이닝 기법을 활용한 사고 심각도 결정 모델 개발	100 %
		A-2-④	정확도 산출 수치를 활용한 기준 모델과의 성능 비교	100 %

2. 관련분야 기여도

가. 기술적 효과

- 연구를 통해 도출된 결과는 인적요인 뿐만 아니라 도로 환경적 요인이 교통사고 발생에 미치는 영향을 심층적으로 보여 줄 것임. 기존에 사용되던 교통사고 분석의 경우, 음주운전이나 부주의한 운전 등, 운전자의 행동 및 과실에만 의존했기 때문에, 교통사고 위험도를 증가시키는 도로 지형적, 환경적 요인들을 감지해내기 어려움. 연구를 통해 얻어낼 분석기술의 경우 다양한 위험요소들을 고려할 뿐만 아니라, 기존 체계보다 정밀한 데이터 마이닝 알고리즘을 적용하여 정부의 교통사고 관련 정책 선정 및 관련 산업체의 업무 효율을 증대시킬 수 있음
- 데이터의 통합 분석을 바탕으로 양적 측면에 집중된 국내 DB의 질적 취약성을 보완할 수 있음. 현재 우리나라의 교통사고 관련 DB의 경우, 경찰, 보험회사, 병원 등 다양한 기관에서 다양한 기준으로 기록되고 있으나, 불성실한 기록 등으로 인해 데이터의 질이 떨어지고, 통합되지 않은 상태로 각 기관에 의해 관리되어 효율적인 이용이 어려움. 본 연구에서 이용되는 데이터베이스 통합 기술을 이용하여, 선진화된 미국의 교통사고 DB 시스템인 SWITRS 와 같은 한국형 교통사고 DB 구축의 발판을 마련할 수 있음
- 본 연구를 통해 개발된 분석모델은 보다 안전한 스마트 도로교통 시스템 개발에 기여할 것으로 기대함

나. 경제적 효과

- 선제적 위험분석을 통한 예방적 차원의 대처가 가능해지면서 물적, 인적, 사회기관비용 감소. 사고 위험요인 분석과정에서 도로환경적 및 상황적 요인을 고려함으로써, 교통사고 위험지역 우선 정비 등에 자원 및 재정을 투자할 수 있고, 후속 지출을 막을 수 있음
- 빅데이터 시장의 지속적인 성장에 따른 교통 분야의 빅데이터 분석기술 수요 증가 및 매출 증대를 기대할 수 있음
- 교통사고 분석기술을 포함한 개발기술의 해외 시장 진출이 가능함

다. 사회적 효과

- 보다 안전한 도로교통 시스템을 통한 국민 체감적 안전 증진 및 선진형 도로 교통문화 조성에 크게 이바지 할 것으로 예상함
- 교통사고 위험분석을 통해 파악된 주요 위험요인을 바탕으로 교통안전증진을 위한 정책적 제안이나 효율적인 사고예방과 위험경감 조치 방안을 모색하는데 핵심적인 역할을 할 것으로 기대함



Copyright © TRUE lab, Korea Advanced Institute of Science and Technology

그림 53. 최종성과물의 기대효과 및 파급효과

제5장. 연구개발성과의 활용계획

- 정부의 교통 정책 분야에 있어서 본 연구는 정밀한 위험요인 분석을 통하여 신뢰성 있는 정보를 제공함으로써 효과적인 교통안전 정책 수립과 교통 행정 처리를 도모할 수 있음
 - 교통사고 위험지역 선정기준 개선, 교통사고 유형별 대응 솔루션 개발, 정비 대상 도로 선정 등의 과정에서 신뢰도 높은 자료를 제공할 것으로 예상함
- GIS 기반으로 한 다른 데이터베이스 기술과의 연계 활용이 가능할 것으로 예상됨. 그 예로서, 교통 흐름 및 사고 데이터를 기반으로 한 경로 안내 체계 및 교통 상황 예측 기술 개발 등에 연계되어 제공된 교통사고 위험요소에 대한 특성 및 정보가 활용 가능할 것으로 기대함
- 이번 연구를 통하여 구축된 통합 교통 DB는 우수한 질적 수준을 바탕으로 여러 관점에서 지속적인 활용이 가능한 연구적 가치를 지님
 - 친환경 교통수단으로 그 수요 및 이용량이 꾸준히 증가하고 있는 자전거에 관한 교통안전 연구
 - 유가 상승과 더불어 수요가 증가하고 있는 택시, 버스 등 대중교통 연관 교통사고에 대한 연구
- 앞서 언급했듯, 14세 이하 어린이 인구 10만명당 보행 사망자수가 0.7명에 달해 OECD 회원국 평균(0.4명)에 비해 두 배 가까이 많고, 65세 이상 노인 인구 10만명당 보행 사망자수는 15.6명으로 1위를 기록하며 OECD 회원국 평균(3.3명)과 비교할 때 무려 5배에 달하는 등, 심각한 수준의 교통약자인 어린이와 노인층 보행 교통사고에 대한 연구가 가능해짐으로서, 고령화 사회에 따른 사회적 현안에 대한 대안 제시가 가능해 질 것임
- 뿐만 아니라, 데이터베이스 통합 기술 역시 향후 교통사고 연관 후속연구에 응용될 수 있는 다양한 분야와의 협력연구에 사용될 수 있음
 - 자동차 보험회사 데이터와 연결하여 교통사고 유형별 피해규모 예측모델 개발 연구
- ITS 기술의 발달로 인한 스마트 도로 교통 시스템에서 선제적 교통안전 구현을 위한 핵심기술로 활용 가능함. 최근 많은 연구를 통해 개발되는 다양한 교통정보 전달 장치를 통해 실시간으로 도로교통 환경 및 위험요인 분석 결과를 바탕으로 한 도로의 위험 정보를 운전자 및 도로 관리자에게 제공함으로써 사전적 교통사고 예방 기술을 구현할 수 있을 것으로 예상함

-
- 실시간 도로 상황 데이터를 사용하여 운전자에게 예상 교통흐름 변화 안내, 교통사고 위험요인에 대한 위험성 알림 등의 역할을 수행하는 내비게이션 형태의 어플리케이션 개발에 응용이 가능함
 - 개인용 어플리케이션뿐만 아니라, 교통방송 등에서도 교통흐름 변화 안내를 위한 핵심기술로 사용될 수 있음

제6장. 연구 과정에서 수집한 해외 과학기술 정보

1. 관련분야의 국내외 연구동향 및 학술적 흐름 분석

- 교통사고 이력 데이터를 분석하는 기법에는 기존의 회귀 분석을 활용한 통계적 기법과, 새롭게 각광받고 있는 데이터 마이닝 기법으로 나뉨. 통계적인 방법으로 처리하는 회귀 분석의 경우, 매개 기법으로 각 요인들 간의 상관관계가 없어야함이 전제되어야 하기에 서로 상관관계가 존재할 수 있는 많은 요인들에 적용하기 어려운 점과, 회귀 분석을 하기 위한 선행 수식의 형태를 먼저 선정해야하는 어려움이 있음.
- 기존에 가장 많이 활용된 확률 분포는 푸아송 분포와 음이항 분포로 실제 현상의 특성과 가장 유사한 성질을 지닌 확률 분포를 사용해왔음. 그러나 데이터 마이닝 기법의 경우 비매개 기법으로 요인들 간의 상관관계가 전혀 문제가 되지 않으며, 그러면서도 요인들 간의 상관관계를 밝혀주는 역할을 할 수 있음.
- 2000년대 중반에 이르러 데이터 마이닝 기법을 활용한 교통사고 이력 데이터 활용들이 활발하게 연구되기 시작했으나, 현재 우리나라의 관련 연구 성과는 한 건 정도로 미비한 실정임.

2. 국외 관련 기술 특허 동향

- 일반적으로 교통사고, 교통안전 분야의 특허는 도로시설이나 자동차 등을 통한 시스템의 교통사고 자동 인식, 자료수집 및 인터넷이나 단말기를 통한 실시간 도로 상황 안내에 치중하고 있고, 미국과 유럽도 비슷한 실정임.
- 연관성은 비교적 낮지만 특기할만한 선행 특허는 1건으로 “교통사고 데이터 제공 시스템 및 방법 System and Method for Providing Traffic Accident Data”로서 본 발명은 교통사고 데이터 제공 시스템에 관한 것으로, 제1 사용자 단말로부터 교통사고가 발생된 현장에 대한 교통사고 영상 정보와 교통사고 위치 정보를 포함한 교통사고 데이터가 전송되면, 교통사고 데이터를 수집하여 저장하는 데이터 수집부 및 저장된 교통사고 데이터를 분석하여 사고 다발 지역을 선정하는 데이터 분석부를 포함하는 사고 관리 서버를 포함하는 내용의 특허임
- 그러나 이 기술은 사고다발지역의 실시간 집계 선정 및 제공을 위하여 교통사고 데이터를 분석하여 정해진 지역에 대하여 교통사고 발생건수를 산출하는 것이기에 사고 심각도 모델을 바탕으로 하여 위험 요인을 분석하는 모델을 개발하는 본 과제와 연관성 및 분쟁 가능성은 매우 낮은 것으로 판단됨

제7장. 연구개발성과의 보안등급

가. 보안등급의 분류

보안	일반
	○

나. 결정사유

국가연구개발사업 공통 보안관리 규칙(교육과학기술부령) 및 국토교통기술연구개발사업 보안관리지침(국토교통부 훈령)에 따라 본 과제는 일반등급의 보안등급으로 분류됨
--

제8장.국가과학기술종합정보시스템에 등록한 연구시설·장비 현황

구입 기관	연구시설/연구장비명	규격 (모델명)	수량	구입 연월일	구입 가격 (천원)	구입처 (전화번호)	비고 (설치 장소)	NTIS장비 등록 번호
-	-	-	-	-	-	-	-	-

제9장. 연구개발과제 수행에 따른 연구실 등의 안전 조치 이행 실적

해당사항 없음

제10장. 연구개발과제의 대표적 연구 실적

☐ 논문(국내외 전문 학술지) 게재

번호	논문명	학술지명	주저자명	호	국명	발행기관	SCI 여부 (SCI/비SCI)	게재일
1	Application of classification algorithms for analysis of road safety risk factor dependencies	Accident Analysis and Prevention	Oh-hoon Kwon	75		Elsevier	SCI	2014.11.16

☐ 국내 및 국제 학술회의 발표

번호	회의 명칭	발표제목	발표 일시	국명
1	대한교통학회 72회 학술발표회	교통정보 및 지리정보체계를 반영한 구급차 출동범위 분석	2015.02.14	대한민국
2	2016 INFORMS International Conference - Hawaii	Comparative Analysis of Various Population Groups in Car Collision Based on Collision Risk Factors Using Data Mining Technique	2016.06.12	미국
3	2016 INFORMS International Conference - Hawaii	Identifying Range Anxiety of Electric Vehicle Drivers Using Psychophysiological Measures	2016.06.12	미국
4	대한교통학회 75회 학술발표회	서울시 어린이 보행자 사고 위험요인 분석에 관한 연구	2016.09.23	대한민국

☐ 전문 연구 인력 양성

번호	분류	기준 연도	현황										
			학위별				성별		지역별				
		2014.08-2016.08	박사	석사	학사	기타	남	여	수도권	충청권	영남권	호남권	기타
				4			3	1		4			

제11장. 기타 사항(성과증빙)

1. 성과증빙 - 논문

가. 논문 게재 - Accident Analysis and Prevention



Application of classification algorithms for analysis of road safety risk factor dependencies



Oh Hoon Kwon^a, Wonjong Rhee^b, **Yoonjin Yoon^{a,*}**

^a Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daejeon-ro, Yuseong-gu, Daejeon 305-701, South Korea

^b Graduate School of Convergence Science and Technology, Seoul National University, 145 Gwanakgyo-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-270, South Korea

ARTICLE INFO

Article history:

Received 18 July 2014

Received in revised form 13 October 2014

Accepted 1 November 2014

Available online 16 November 2014

Keywords:

Traffic accident analysis

Severity level

Risk factor

Naive Bayes classifier

Decision tree classifier

Dependency

ABSTRACT

Transportation continues to be an integral part of modern life, and the importance of road traffic safety cannot be overstated. Consequently, recent road traffic safety studies have focused on analysis of risk factors that impact fatality and injury level (severity) of traffic accidents. While some of the risk factors, such as drug use and drinking, are widely known to affect severity, an accurate modeling of their influences is still an open research topic. Furthermore, there are innumerable risk factors that are waiting to be discovered or analyzed. A promising approach is to investigate historical traffic accident data that have been collected in the past decades. This study inspects traffic accident reports that have been accumulated by the California Highway Patrol (CHP) since 1973 for which each accident report contains around 100 data fields. Among them, we investigate 25 fields between 2004 and 2010 that are most relevant to car accidents. Using two classification methods, the Naive Bayes classifier and the decision tree classifier, the relative importance of the data fields, i.e., risk factors, is revealed with respect to the resulting severity level. Performances of the classifiers are compared to each other and a binary logistic regression model is used as the basis for the comparisons. Some of the high-ranking risk factors are found to be strongly dependent on each other, and their incremental gains on estimating or modeling severity level are evaluated quantitatively. The analysis shows that only a handful of the risk factors in the data dominate the severity level and that dependency among the top risk factors is an imperative trait to consider for an accurate analysis.

© 2014 Elsevier Ltd. All rights reserved.

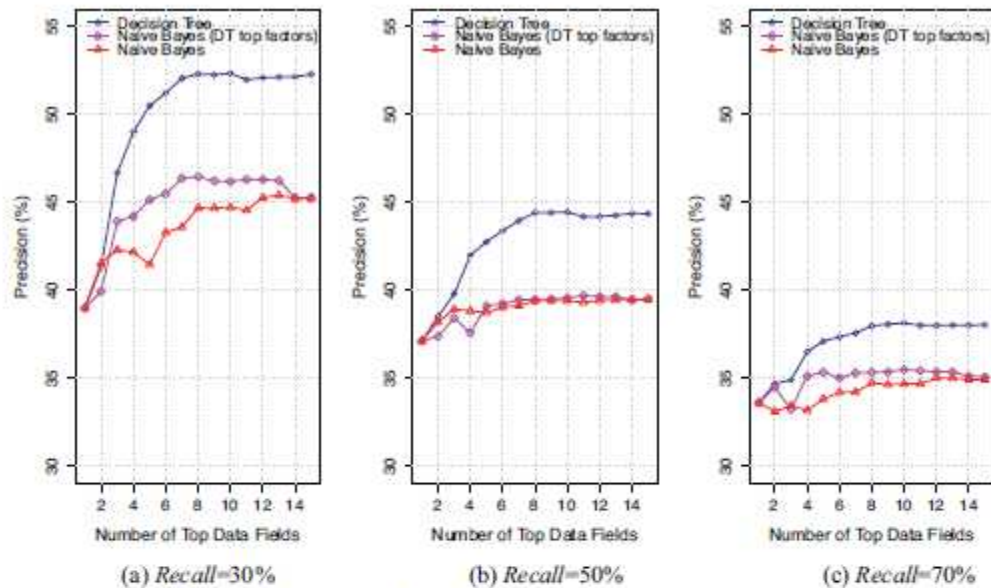


Fig. 7. Precision improvements for fixed recall according to the number of top-ranked risk factors.

7. Conclusions

A variety of traffic accident data sets has been studied in the past decade, and the most important risk factors have been identified for the associated region and time period of each data set. In this study, the dependency of risk factors was analyzed for California data, and some of the top risk factors were found to carry similar, and thus redundant, information. Two classification methods and two related scoring algorithms were used for finding rankings of the risk factors, where the first method, Naive Bayes, does not reflect dependency but the second method, the decision tree, does. By comparing a wide range of precision and recall performances of the two classifier approaches and logistic regression model, we found that two out of five top risk factors from Naive Bayes are moved outside of the top five when dependency is considered. Furthermore, the decision tree, which accounts for dependency, always outperformed the Naive Bayes and the logistic regression models. To confirm that the performance gap is, at least partly, due to the consideration of dependency, Naive Bayes was re-evaluated with its own top five risk factors only and then with the decision tree's top five only. The comparison showed that the decision tree's choice of the top five risk factors is superior in terms of allowing Naive Bayes to perform better and that the ranking of risk factors from the decision tree classifier could contribute to efficient safety countermeasures in limited resources. This result indicates that it is important to consider dependency among the risk factors and to adopt appropriate analysis methods such that the risk factors are fairly and accurately evaluated. The methodologies used in this paper are readily applicable to the other existing data sets for evaluating dependency among risk factors and for understanding their relative importance on dictating severity.

Data fields and possible values of accident data vary with government agency who collects the data. If there were some missing important risk factors in the collection, the performance of the Naive Bayes classifier could decrease due to the loss of the important information. On the other hand, the decision tree classifier could find alternative risk factors considering the dependency. For this reason, the decision tree classifier could be less sensitive to input risk factors than the Naive Bayes classifier in the performance. This sensitivity analysis needs to be conducted clearly as future study.

For the California data, five to eight risk factors were found to dominate the severity of traffic accidents. Without a proper consideration of dependency, we found the following factors to be the most important: collision type, violation category, movement preceding collision, intersection, and state highway. With a proper consideration of dependency, we found these factors to be the most important: collision type, at fault, population, state highway, and movement preceding collision. Note that the decision tree that addresses dependency replaces two out of the five top risk factors.

Acknowledgements

This research was supported by Ministry of Land, Infrastructure and Transport grant 14CTAP-C078824-01, and Brain Korea Plus research and education program 21A20132000003. Authors would like to thank the reviewers for very helpful comments.

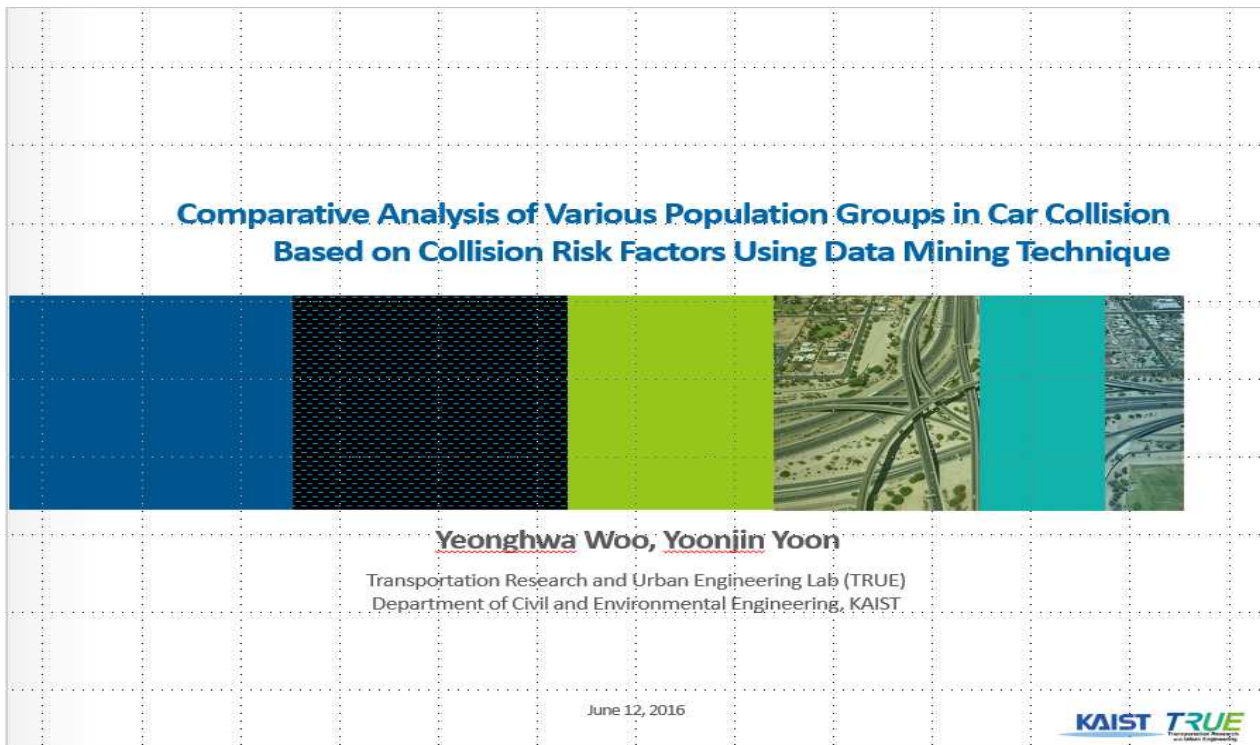
Appendix A.

Table A1 Distribution of severity per data field.

나. 학술대회

(1) 국외 학술대회 - 2016 INFORMS International Conference - Hawaii

(가) Comparative Analysis of Various Population Groups in Car Collision Based on Collision Risk Factors Using Data Mining Technique

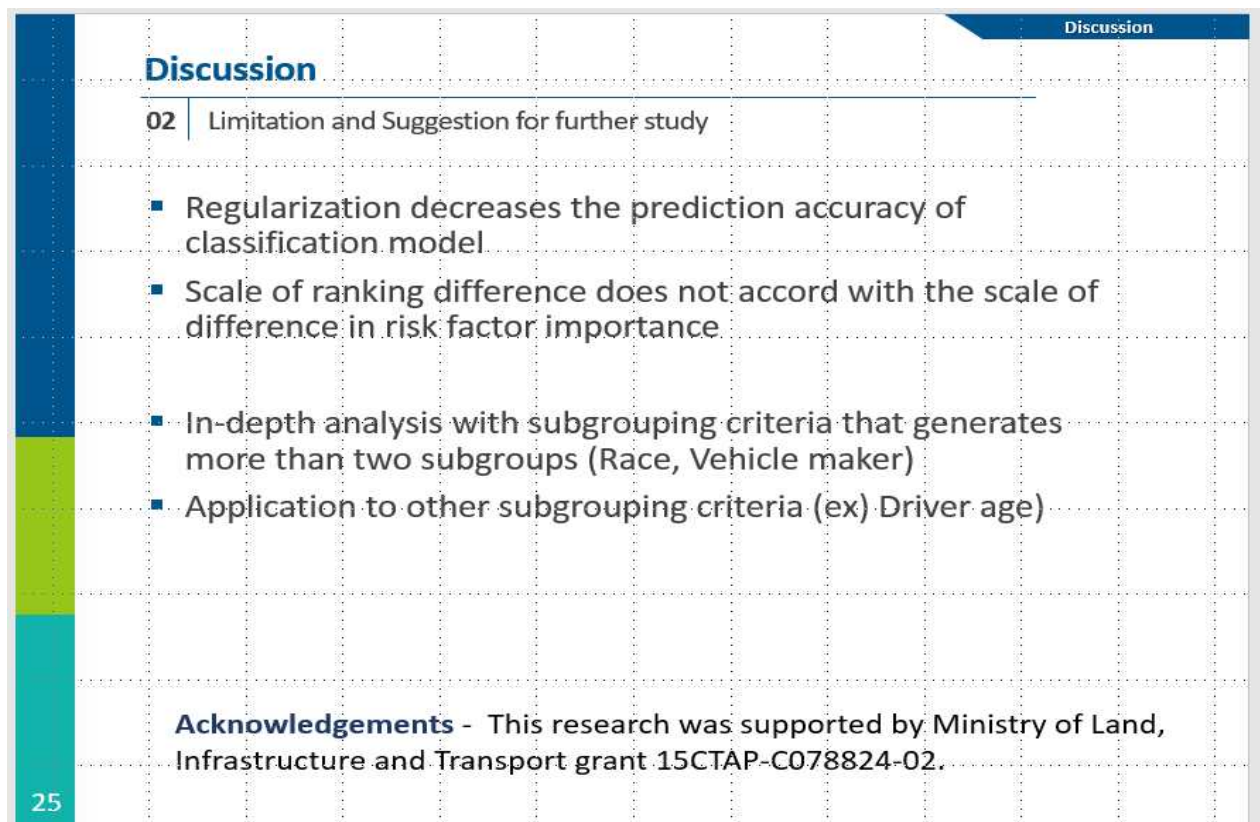


**Comparative Analysis of Various Population Groups in Car Collision
Based on Collision Risk Factors Using Data Mining Technique**

Yeonghwa Woo, Yoonjin Yoon
Transportation Research and Urban Engineering Lab (TRUE)
Department of Civil and Environmental Engineering, KAIST

June 12, 2016

KAIST TRUE
Transportation Research and Urban Engineering



Discussion

02 | Limitation and Suggestion for further study

- Regularization decreases the prediction accuracy of classification model
- Scale of ranking difference does not accord with the scale of difference in risk factor importance
- In-depth analysis with subgrouping criteria that generates more than two subgroups (Race, Vehicle maker)
- Application to other subgrouping criteria (ex) Driver age

Acknowledgements - This research was supported by Ministry of Land, Infrastructure and Transport grant 15CTAP-C078824-02.

25

Identifying Range Anxiety of Electric Vehicle Drivers Using Psychophysiological Measures

Dongwoo Jin, Seyun Kim, Yoonjin Yoon

Department of Civil and Environment Engineering, KAIST

2016. 06. 12

INFROMS 2016

KAIST TRUE 



Conclusion

► Contribution

- The first attempt of identifying range anxiety by analyzing psychophysiological data collected from human beings.
- Can be utilized for developing HVI (Human Vehicle Interface) by customized information for safe driving.

► Limitation

- All test trial started with similar SOC level.
- Subjects participated the test only once.
- Recording environment of biosignal sensors.

► Future Direction

- Advanced experimental design (critical situation, tracing subjects)
- Cooperating with experts for processing biosignal data
- Comparing to other SOC researches (cellphone)

Acknowledgements - This research was supported by Ministry of Land, Infrastructure and Transport grant 15CTAP-C078824-02.

KAIST TRUE 21

3 - Local Energy Storage As Part of Homeland Security

Dean S. Hartley, Hartley Consulting, DSHartley3@comcast.net

We discuss the security benefits of breaking the national energy grid into a thousand pieces. We show how local energy storage supports this and how the Skyphus system enables local energy storage.

SA03

King's 3

Behavioral Operations and Cars

Sponsored: Behavioral Operations Management

Sponsored Session

Chair: Natalia Sowik, EBS University of Business and Law,
Gustav-Stresemann-Ring 3, Wiesbaden, 65189, Germany,
natalia.sowik@ebs.edu

1 - Design of Pedestrians Running Red Lights Warning System

Changan Gao, Huazhong University of Science and Technology,
Hongshan District, 1037 Luo Yu Road, Wuhan, 430074, China,
bestchangan@126.com

In China, many pedestrians may run a red light when crossing the road. In order to solve this problem, we conducted a field survey, and collected a large number of data. Then we analyzed the reason of the pedestrian red light, and designed an alarm device based on the knowledge of electromagnetism, optics, acoustics and so on. This device and the traffic light system can be combined into an intelligent alarm system, whose application in the intersection can effectively improve the phenomenon of pedestrians running red lights, and help to improve the traffic efficiency.

2 - Psychological Ownership In Access Based Consumption on the Example of Carsharing

Natalia Sowik, EBS University of Business and Law, Gustav-
Stresemann-Ring 3, Wiesbaden, 65189, Germany,
natalia.sowik@ebs.edu

Humans express themselves through ownership (OS) by building consumer-object relationships. Nevertheless, access-based consumption (ABC), formed on the avoidance of OS, but acquisition of temporary access to objects, is gaining momentum. Based on the shift from OS to access, we assume that it is not essential to own an object to develop feelings of OS, referring to the theory of psychological ownership (PO). To examine PO in ABC, we conducted a quantitative study among 1,526 carsharing users. The results show evidence for PO with no significant difference between respondents, who additionally own a car, and those, who do not. Impacts of PO on satisfaction, loyalty, and word-of-mouth were found.

3 - Identifying Range Anxiety of Electric Vehicle Drivers using Psychophysiological Measures

Dongwoo Jin, KAIST, yuseong-gu Daehak-ro 291, W1-2 4203,
Daejeon, Korea, Republic of, jindwoo@kaist.ac.kr

Electric vehicles were developed as the alternative eco-friendly transportation to reduce dependency on fossil fuel. However, technical and psychological difficulties from short driving range, slow recharging speed and fewer recharge stations still limit the broad acceptance of electric vehicles in the public. This study focused on the phenomenon of range anxiety which is drivers' anxiety about the limited range of their electric vehicles and underdeveloped infrastructure of charging opportunities. EEG, ECG and pupillary size were utilized for this study and they first processed to extract anxiety-related indexes for representing drivers' mental status while driving electric vehicles.

4 - Comparative Analysis of Various Population Groups in Car Collision Based on Collision Risk Factors using Data Mining Technique

Yeonghwa Woo, KAIST, Daejeon, Korea, Republic of,
beeluebam@kaist.ac.kr

Road safety is one of main health issues related to transportation system. Risk factor analysis is performed to identify major variables that determine severity of collisions, and to measure their influence to the injury severity. In this study, we identify major risk factors that determine collision severity for various subgroups involved in accidents, and compare different subgroups quantitatively and qualitatively based on risk factor importance. Subgroups are generated by five criteria of driver's gender, population of collision area, at-fault party indicator, race of driver, and vehicle maker. Risk factor importance is measured using data mining technique of regularized random forest.

SA04

Queen's 4

Marketing and Commerce

Contributed Session

Chair: Osman Oguz, Bilkent University, Dept of Industrial Engineering,
Ankara, 06800, Turkey, ooguz@bilkent.edu.tr

1 - The Impact of Business Response to Negative Online Reviews

Shuoqia Guo, Assistant Professor, CUNY-CSI, 2800 Victory
Boulevard, Staten Island, NY, 10314, United States,
nancy.guo@csi.cuny.edu

Online feedback mechanisms have become increasingly important to companies' bottom lines as more and more consumers rely on online reviews to evaluate product or service quality before making purchase decisions. However, it is almost certain that at some point some consumers will write negative reviews for a company. While most of the extant literature on online reviews is focused on how the valence of online reviews affects consumers' purchase intentions, there has been virtually no attention paid in the topic of online business responses. This current study aims to investigate how managers' responses to customer reviews and in particular to negative reviews affect subsequent ratings and reviews.

2 - Exploring the Biasing Effects of a Two-Sided Format on Product Reviews

Haksin Chan, Hong Seng Management College, Shatin,
Hong Kong, hshan@hsmc.edu.hk

Online product reviews are among the most frequently used information sources today. Interestingly, online review forums vary widely in how they solicit information. Some forums give reviewers a free hand in the review process, whereas others request reviewers to highlight both positive and negative product attributes. Intuitively, the two-sided format (TSF) is conducive to unbiased reviews. This research, however, explores the potentially biasing effects of the TSF. Results from three experiments indicate that the TSF can have a centralizing or a polarizing effect on product ratings. These results have strong implications for understanding, comparing, and driving user-generated content.

3 - Mobile Shopping More Impulse Purchases in a Higher Friction Market

Yubo Chen, Associate Dean and Professor, Tsinghua University,
School of Economics and Management, Beijing, 100084, China,
chenyubo@sem.tsinghua.edu.cn

The rapid development of the mobile Internet is changing the landscape of the E-commerce. It becomes very important to understand whether and how the mobile channel affects consumer purchase behavior. This paper studies this issue by modeling the data from one of the largest online retailers in China. We find that customers' mobile channel adoption can significantly increase their overall purchase frequency and spending in both the short and long term. Interestingly, the mobile channel can have both substitutive and complementary effects on the traditional PC channel. We further examine the mechanism behind the impacts of mobile channel on customer purchase behaviors.

4 - A Heuristic for Travelling Salesperson Problem

Osman Oguz, Assoc. Professor, Bilkent University,
Dept of Industrial Engineering, Ankara, 06800, Turkey,
ooguz@bilkent.edu.tr

A Dynamic Programming based heuristic algorithm is presented to solve the TSP. The complexity of the algorithm is polynomial. The performance of the algorithm will be evaluated on a set of numerical experiments.

(2) 국내 학술대회

(가) 대한교통학회 72회 학술발표회 - 교통정보 및 지리정보체계를 반영한 구급차 출동범위 분석



KAIST

제72회 학술발표회 (2015년도 춘계학술대회)

교통정보 및 지리정보체계를 반영한 구급차량 출동 범위 분석

조정우 · 윤운진
한국과학기술원 건설및환경공학과



TRUE
Transportation Research
and Urban Engineering

· 학술지도자: [주9] 71 [주9] 71 [주9] 71, Phone: +82-10-5576-7519

· 본 연구는 국토교통부 국토교통기술촉진연구사업의 연구비 지원(과제번호: 15CTAP-C078824-01)에 의해 수행되었습니다.

연구의 배경 및 목적

- 응급의료체계에서 긴급차량의 신속한 현장 도착 및 구급처치는 사고 피해자의 생존과 매우 밀접한 관련이 있음 [1][2]
- 한국을 비롯한 주요 국가에서 사고현장에서 초동 조치에 영향을 미치는 긴급차량의 현장 도착시간 개선은 지역 여건과 같은 목표를 정함
- 미국의 Denver에서는 최소 90%의 긴급상황에 대해 긴급차량이 반응시간(신고발령-현장도착) 8분내에 현장에 도착하도록 목표를 설정함 [3]
- 캐나다의 경우 90%의 긴급전화를 반응시간 9분내로 처리하는 것을 목표로 함
- 영국에서는 75%의 긴급상황에 대해 긴급차량이 신고발령시부터 8분내에 현장 도착하는 것을 목표로 함 [4]
- 한국은 2017년까지 74%의 긴급상황에 대해 긴급차량의 119센터 출동부터 사고현장 도착까지 5분내에 완수하도록 목표를 설정함 [5]

구급차 유형	구급차 출동	구급차 도착	구급차 현장도착	구급차 현장도착	구급차 현장도착
응급구급차	119	119	119	119	119

그림 1. 구급차 이동시간 통계, 평균(표준편차)

결과

연구목표: GIS와 속도데이터를 사용하여 서울시 119 안전센터의 5분 내 도달가능 범위의 시간대별 변화량 분석

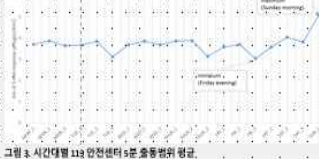


그림 3. 시간대별 119 안전센터 5분 출동범위 평균

- 과거 교통 데이터, 서울시 도로망과 소말리아 지리정보를 반영하여 GIS상에서 각 소말리아별 출동 범위의 다양한 시간대에 따른 5분 출동범위를 분석한 결과는 그림 2와 같음
- 평균적으로 차량 통행이 현저한 일요일 아침에 비해 가장 혼잡한 평요일 저녁에 출동 범위가 평균 40% 감소하는 등 상당한 변동폭이 있음
- 특히 119 안전센터의 경우 교통속도 변화에 따라 출동범위가 70%까지 감소함을 알 수 있음
- 즉, 각 요일별 특정 시간대의 교통상황에 따라 출동범위가 유동적으로 변함을 알 수 있고, 119 안전센터의 긴급차량 출동 의사결정에 있어서 교통상황이 중요한 요인임이 증명됨
- 출동범위가 가장 큰 일요일 오전, 가장 작은 평요일 저녁을 기준으로 분석을 진행하였고, 두 시간대에 따른 출동범위 차이는 그림 4와 같음
- 중구, 서구, 강남구, 강서구 등 시간대의 출동범위와 상대요자가 가장 큰 구역으로 밝혀지며, 이는 해당 구역이 교통량이 가장 밀집한 구역을 의미하고, 4개의 구가 강북, 강남의 주요 비즈니스 지구임을 고려했을 때 타당한 결과임 (표3)
- 각 구의 위치에 따라 크게 강북구, 강남구 등으로 나눌 수 있음. 앞으로의 분석에서 두 구를 중심으로 비교할 것임

그림 4. 평요일 저녁과 일요일 오전의 5분 출동범위 차이

표 3. 두 시간대의 119 안전센터 5분 출동범위와 상대요자

구역	119 안전센터 개수	일요일 오전 5분 출동범위, km²	평요일 저녁 5분 출동범위, km²	상대요자
중구	5	4.9	2.4	70%
서구	6	4.8	2.5	69%
강남구	5	3.3	2.9	59%
강북구	6	4.8	2.7	56%
강서구	5	4.8	3.1	43%

교통정보체계

51	정당도로교통 시설물 통합표준제어의 경제성 분석 연구	김동민 (도시연구실 석사과정) 김정석 (공주대학교 교수)	Po-7
52	교통정보 및 지리정보체계를 반영한 구급차 출동 범위 분석	조정우 (한국과학기술원 석사과정)	Po-8

교통정책및계획

59	유성복합터미널 조성사업의 필요성 분석	조영남 (한밭대학교 석사과정) 김명수 (한밭대학교 교수) 유병준 (한밭대학교 학사과정) 신지혜 (한밭대학교 학사과정)	Po-9
60	환동해권 사회간접자본 현황 및 물류 선진화 방안	김길환 (한양대학교 석사과정)	Po-10

The 72th Conference of Korean Society of Transportation (2015. 2. 13-14) 13

52

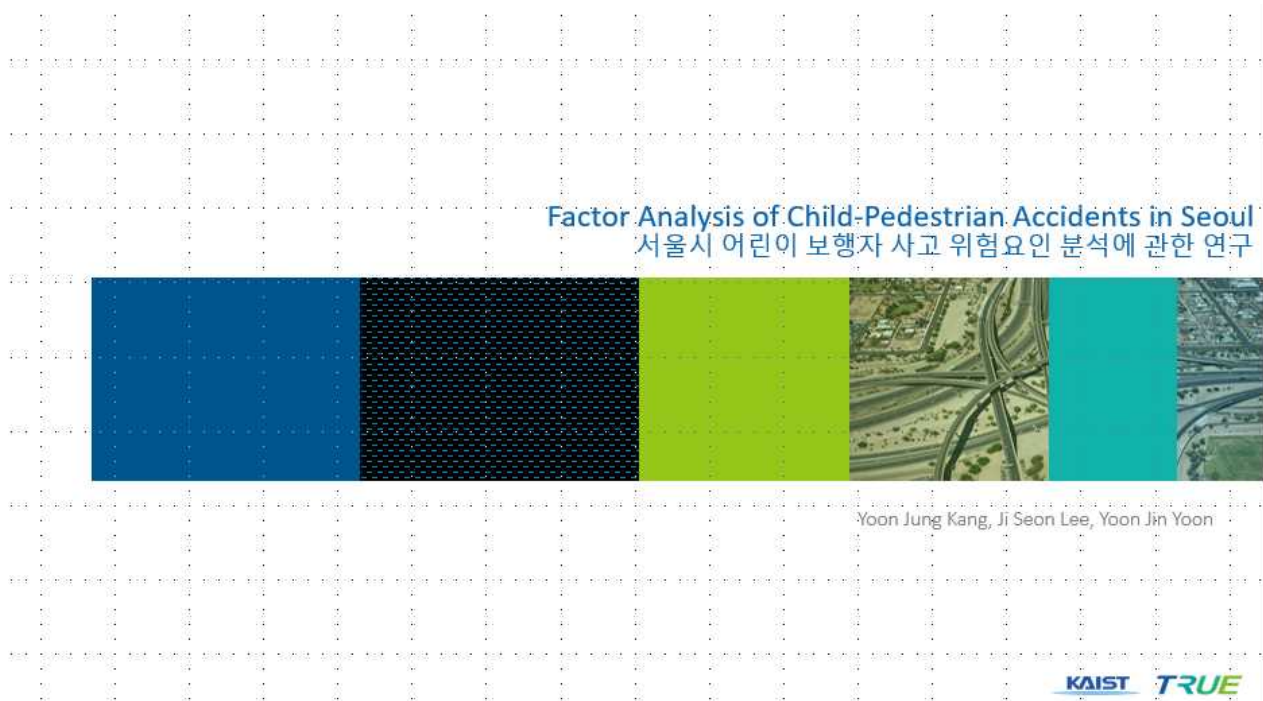
교통정보 및 지리정보체계를 반영한 구급차 출동범위 분석

교통정보 체계

조정우 (한국과학기술원 석사과정)

윤운진 (한국과학기술원 조교수)

응급의료체계에서 긴급차량의 신속한 현장 도착 및 구급처치는 사고 피해자의 생존과 매우 밀접한 관련이 있다. 주요 국가에서 사고현장에서의 초동 조치에 영향을 미치는 긴급차량의 현장 도착시간 개선을 위한 노력을 하고 있으며, 한국에서는 2017년까지 74%의 긴급상황에 대해 119 출동부터 사고현장 도착까지 5분내에 완수하는 것을 목표로 하고 있다. 본 연구에서는 서울시 119안전센터의 출동범위를 서울시의 GIS 도로망 정보와 과거 교통속도 데이터를 바탕으로 분석하고자 한다. 여기서 출동범위는 5분 내로 도달 가능한 지역으로 하고, 각 요일의 다양한 시간대의 교통속도에 따라 유동적으로 변하는 출동범위를 분석하였다.



Conclusion

- Identified children pedestrian road accident characteristics
 - pedestrian accident is about 20% of total traffic accidents
 - serious injury rate (severe or fatal) in pedestrian accident is higher than the overall traffic accident
 - children pedestrian injury occurs mostly during school on/ off time and school months
- Developed pedestrian injury risk map by Seoul district based on pedestrian injury rate and its population
 - normalization of pedestrian accident rate by population in 25 districts
 - Jongnang-gu has the highest children pedestrian injury rate
 - Yanchun-gu has the highest proportion of children pedestrian injury in total pedestrian injury
- Proposed children pedestrian injury prediction model using stepwise regression
 - significant variables are number of students and district financial independence
 - improvement in district financial independence can result in the reduction of children pedestrian injury
- Further Study
 - Include other variables
 - Accurate model development (Time-series)

Acknowledgements - This research was supported by Ministry of Land, Infrastructure and Transport grant 15CTAP-C078824-02.

교통안전 2 23(금) 09:30-11:00, 203		좌장 : 박신형 (계명대학교) 한상진 (한국교통연구원)
22	대형교통사고 특성과 사고 유발인자 분석	정동훈 (도로교통공단 사고조사연구원) 정순도 (도로교통공단 안전본부장) 박길수 (도로교통공단 센터장) 문정식 (도로교통공단 사고분석개선처장)
23	도로요인이 위험운전 유형에 미치는 영향	이상용 (한국교통연구원 주임전문원) 김용철 (인천대학교 교수) 신희철 (한국교통연구원 연구위원)
24	2단계 군집분석을 통한 지형과 교통사고와의 관계 분석: 강원도를 대상으로	곽하영 (경희대학교 석사과정) 백승걸 (한국도로공사 수석연구원) 박운호 (경희대학교 박사과정) 조창현 (경희대학교 부교수)
25	도로환경변화에 따른 보행자 횡단사고 발생 요인에 관한 연구: 대구도시철도 3호선 하부도로를 중심으로	김민경 (계명대학교 석사과정) 강민지 (계명대학교 석사과정) 권오훈 (계명대학교 조교수) 박신형 (계명대학교 조교수)
26	단독주택지와 공동주택지의 교통사고 위험도 비교	이해선 (한국교통연구원 연구위원) 한상진 (한국교통연구원 연구위원) 이해지 (한국교통연구원 연구위원)
27	서울시 어린이 보행자 사고 위험요인 분석에 관한 연구	강윤정 (한국과학기술원 석사과정) 이지선 (한국과학기술원 박사과정) 윤윤진 (한국과학기술원 조교수)
교통안전 3 23(금) 09:30-11:00, 204		좌장 : 이범규 (대전발전연구원) 이백진 (국토연구원)
28	도로안전 관련 시설물 교통사고 감소효과도 추정	윤여일 (도로교통공단 대리) 정순도 (도로교통공단 안전본부장) 박길수 (도로교통공단 센터장) 문정식 (도로교통공단 사고분석개선처장)
29	경험적 베이지 방법을 이용한 구간과속단속시스템 설치효과 분석	조준호 (명지대학교 연구원) 이의은 (명지대학교 교수) 박병정 (명지대학교 조교수) 신형오 (명지대학교 박사과정) 정가형 (명지대학교 석사과정)
30	4Q Analysis를 이용한 국가간 교통사고 수준 비교	엄기종 (한국교통연구원 연구위원) 한상진 (한국교통연구원 연구위원)
31	교통사고와 교통지도단속의 상관분석 연구	김주영 (교통안전공단 책임연구원) 김현진 (교통안전공단 책임연구원) 손나영 (교통안전공단 선임연구원)
32	사고위험도 분석을 통한 고령운전자 안전정책 연구	최재성 (국토연구원 책임연구원) 김준기 (국토연구원 연구위원)
33	교통안전측면에서의 업무용자동차의 문제점 및 개선방안	원동욱 ((유)아름엔지니어링 이사)
교통운영 1 23(금) 09:30-11:00, 206-1		좌장 : 김원호 (서울연구원) 신희철 (한국교통연구원)
51	PLS회귀분석을 이용한 MIL-HDBK-217F 환경팩터 값 추정	유지선 (아주대학교 연구원) 오재윤 (아주대학교 연구원) 장중순 (아주대학교 전임교원)
52	링크내 소규모 교차로 발생 지체 보정을 통한 구간통행시간 산출에 관한 연구	조은희 (서울시립대학교 석사과정) 김영찬 (서울시립대학교 교수) 이민형 (서울시립대학교 박사과정)
53	고속도로 면도2차로 장기공사구간의 주야간 용량 변화에 관한 연구	김도훈 (계명대학교 석사과정) 박상화 (계명대학교 석사과정) 권오훈 (계명대학교 조교수) 박신형 (계명대학교 조교수)

제12장. 참고 문헌

- 국내 교통사고 발생현황, 나라지표 - www.index.go.kr
- Road Safety Annual Report 2013, IRTAD
- 교통안전공단 (2014년 5월 8일). “우리나라 보행자 교통사고 OECD 최하위”. 교통안전공단 교통안전처.
- 국토교통부 (2014.2). 2014년 국토교통부 핵심과제 실천계획.
- 서울특별시 (2013년 4월 16일). “서울시, KT와 손잡고 빅데이터로 시민의 작은 고충까지 살피”. 서울특별시 정보기획단.
- 서울특별시 (2014년 4월 22일). “서울시, “수요가 있는 곳에 정책“ 빅데이터 본격 활용”. 서울특별시 정보기획단.
- 박현아 (2014.1). 빅데이터시장 현황과 콘텐츠산업 분야에 대한 시사점. 한국콘텐츠진흥원. 코카포커스 2013-11호.
- 윤미영 (2013.9). 주요국의 빅데이터 추진전략 분석 및 시사점. 과학기술정책연구원. 제23권 제3호.
- 한국방송통신전파진흥원 (2012.7). 빅데이터 처리기술 현황 및 전망. KOMPASS Market&ISSUE.
- IDC (2012.3). Worldwide Big data Technology and Services 2012-2015 Forecast.
- 국토교통부 (2013), “교통정보 활용촉진을 통한 ITS산업 활성화“ 창조산업 육성에 눈트인 국토부, 11월 20일, 국토교통부 보도자료
- 성지수(2014), '똑'소리 나는 도로의 비결? 중국 ITS시장 들여다보기, 1월 24일, KOTRA 해외비즈니스정보포털 globalwindow 해외시장정보
- Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention*, 34(6), 729-741.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press.
- Chang, L. Y., & Chen, W. C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 36(4), 365-375.
- Chang, L. Y., & Wang, H. W. (2006). Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention*, 38(5), 1019-1027.
- Deng, H. (2013). Guided random forest in the RRF package. *arXiv preprint arXiv:1306.0237*.
- Deng, H., & Runger, G. (2012, June). Feature selection via regularized trees. In *Neural*

- Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition*, 46(12), 3483-3489.
- Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. *Safety science*, 49(10), 1314-1320.
- Kim, K., Li, L., Richardson, J., & Nitz, L. (1998). Drivers at fault: influences of age, sex, and vehicle type. *Journal of Safety Research*, 29(3), 171-179.
- Krishnaveni, S., & Hemalatha, M. (2011). A Perspective Analysis of Traffic Accident using Data Mining Techniques. *International Journal of Computer Applications*, 23(7), 40-48. doi:10.5120/2896-3788
- Kwon, O. H., Rhee, W., & Yoon, Y. (2015). Application of classification algorithms for analysis of road safety risk factor dependencies. *Accident Analysis & Prevention*, 75, 1-15
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2(3), 18--22.
- Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accident Analysis & Prevention*, 49, 58-72.
- Morgan, A., & Mannering, F. L. (2011). The effects of road-surface conditions, age, and gender on driver-injury severities. *Accident Analysis & Prevention*, 43(5), 1852-1863.
- Moskal, A., Martin, J. L., & Laumon, B. (2012). Risk factors for injury accidents among moped and motorcycle riders. *Accident Analysis & Prevention*, 49, 5-11.
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention*, 43(5), 1666-1676.
- Shankar, V., Mannering, F., & Barfield, W. (1996). Statistical analysis of accident severity on rural freeways. *Accident Analysis & Prevention*, 28(3), 391-401.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8, 25. doi:10.1186/1471-2105-8-25
- Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. (2004). World report on road traffic injury prevention. In *World report on road*

traffic injury prevention. OMS.

- Yan, X., Radwan, E., & Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis & Prevention*, 37(6), 983-995.
 - Yau, K. K. (2004). Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis & Prevention*, 36(3), 333-340.
 - Caltrans PeMS website (<http://pems.dot.ca.gov/>)
 - TAAS website (taas.koroad.or.kr)
-

주 의

1. 이 최종보고서는 국토교통부에서 시행한 국토교통기술촉진연구사업의 연구보고서입니다.
2. 이 최종보고서 내용을 발표하는 때에는 반드시 국토교통부에서 시행한 사업의 연구개발성과임을 밝혀야 합니다.
3. 국가과학기술 기밀 유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안 됩니다.