

의사결정나무와 시공간 시각화를 통한 서울시 교통사고 심각도 분석

강영옥 · 손세린 · 조나혜
이화여자대학교 사회과교육과

2017.10.26

1. 연구의 배경 및 목적

2. 연구대상, 범위 및 연구방법

3. 주요연구결과

4. 결론 및 정책제언

1. 연구배경 및 목적

연구배경

- ▶ 경제성장과 함께 자동차 보유 일상화 → 자동차 수의 증가 → 교통사고 발생건수 증가
- ▶ 정부차원의 법제도 개선, 공익광고, 도로정비 등으로 교통사고 예방에 힘써 교통사고 발생건수 감소
- ▶ 그러나 최근 5년간 교통사고 건수는 2013년 감소하다가 다시 증가 추세
- ▶ OECD국가 중 자동차 1만대당 교통사고 사망자수 3위, 교통사고 사망자 수는 OECD평균의 2배 (2017, 도로교통공단)

1. 연구배경 및 목적

교통사고의 이해

▶ 인적 요인 :

- 신체, 생리심리, 적성 습관, 태도요인 : 운전자 또는 보행자의 신체적 조건, 위험의 인지와 회피에 대한 판단, 심리적 조건 등
- 운전자의 적성과 자질, 운전습관, 내적 태도

▶ 차량적 요인 :

- 차량구조 장치, 부속품 또는 적하에 관계된 사항

▶ 도로환경적 요인 :

- 도로구조 : 도로의 선형, 노면, 차선수, 경사 등
- 안전시설 : 신호기, 도로표시, 방어책 등
- 자연환경 & 교통환경: 기후, 명암 등 자연조건, 차량통행량, 통행차량 구성, 보행자 교통량 등 교통상황
- 사회환경 : 교통경찰관, 보행자 등의 교통, 도덕 등의 환경구조, 교통정책 및 행정법칙요인, 교통단속과 형사 처벌 등의 요인
- 구조환경 : 교통여건변화, 노선버스 운행의 비합리성, 차량점검 및 정비관리자와 운전자의 책임한계, 차량보안기준 위배, 노후차량, 불량판매 등



▶ 공간적 특성

▶ 시간적 특성

1. 연구배경 및 목적

교통사고 관련 연구

▶ 교통사고 발생요인

- 운전시 논리적인 사고를 하지 않아 의사결정단계에서 결함이 발생하는 경우 운전자의 실수로 이어져 사고 발생할 가능성이 높게 나타남 (Parker, Dianne, et al., 1995)
- 운전자들이 자신의 운전 역량과 사고 위험에 대해 지나치게 낙관적인 판단을 내리는 경향, 여성이 더 낙관적임 (DeJoy, DM., 1992)
- 16~19세 및 75세 이상 운전자의 경우 과속경향, 치명적인 충돌 사건은 여성보다 남성이 더 높음, 여성은 사고로 부상을 입을 확률이 높음 (Massie, D. L, et al., 1995)
- 유류판매량, 통행발생량, 도로의 연장이 높은 지역일수록, 그리고 외국인 인구 구성비가 높을수록 사망사고 증가 (박재영·김도경, 2011)
- 대부분의 약물에서는 약물사용과 관련된 사고 위험이 적거나 보통 수준으로 증가하였는데 반해 마약 사용은 일반적인 사고보다 치명적이고 심각한 상해로 이어지는 경향 존재 (Elvik, R., 2013)
- 가스 시추가 많이 이루어져 천공이 크게 나타난 북쪽카운티는 그렇지 않은 카운티에 비해 2010~2012년에 약15~23% 가량 더 높은 차량 충돌률과 2011~2012년 약 61~65% 정도의 트럭 충돌률 경험 (도로주변 공사의 위험성분석 목적; Graham, J, et al., 2015)
- 남성 운전자, 트럭운전자, 새벽운전, 출근시간 등은 졸음운전과 관련된 사고 위험 요인으로 확인되었으나 심각한 사고를 초래하지는 않았으며 야간운전과 졸음운전은 심각한 부상이 발생하는 것으로 나타남 (Zhang, Guangnan, et al. , 2016)

▶ 데이터마이닝 기법을 사용한 교통사고 요인분석

- 최근들어 방대한 데이터에서 유용한 정보와 지식을 밝혀내고자 하는 데이터마이닝 기법들이 교통사고 분석에도 활용됨
- Wahab & Abdel-Aty(2001) : 신호등이 있는 교차로에서 두 개 차량 간 충돌의 심각도를 예측하기 위해 인공신경망 모델 적용
- De Ona, Lopez, Mijalli & Calvo (2011, 2013), Mujalli & de Ona(2011) : 교통사고 심각도를 분석하기 위해 베이지안 네트워크 방법론 사용
- Chang, L. Y., & Chien, J. T., 2013; Chang, L. Y., & Wang, H. W., 2006; Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F., 2011; 2012 : 의사결정나무 분석방법을 이용한 교통사고 분석
- 의사결정나무 분석은 질적 변수나 양적 변수의 자료 분석이 가능하며, 다변량 분석의 기본 가정인 선형성, 정규분포성, 등분산성을 따르지 않아도 되고, 분석 결과가 나무구조로 되어 있어 해석이 용이하다는 점 등으로 교통사고 분석에도 많이 활용됨

1. 연구배경 및 목적

교통사고 관련 연구

▶ 데이터마이닝 기법을 사용한 교통사고 요인분석 : 의사결정나무 분석 활용 예

- Abellan, Lopez & de Ona (2013): 농촌지역 도로에서 교통사고 심각도에 영향을 미치는 요인 분석, 교통사고 심각도에 영향을 미치는 10개의 규칙을 제안
- 박준태 외 (2011): 교통안전시설물 설치가 교통사고 감소에 영향을 미치는 지 여부를 의사결정나무 모형을 이용하여 분석
- 이재명 외 (2009): 전국 4차선 이상 국도에서 발생한 교통사고 데이터를 분석하여 도로 기하구조 요인과 교통사고간의 관계를 분석
- 이문영(2013): 1일 평균 주행거리에 따른 교통사고 영향요인 분석
- 한상진 외(2014): 화물자동차 사고와 비화물자동차 사고요인의 차이를 의사결정나무 모형을 이용하여 분석
- 박용우외 (2016): 고속도로 공사구간이 사고 심각도에 미치는 영향 분석

▶ 교통사고 시각화 관련 연구

- 2차원의 교통사고 클러스터 지역을 분석하기 위해 커널 밀도 추정을 통한 선형, 픽셀단위, 지역 등으로 시각화한 연구가 진행
- Xie, Z. 외 1명(2008): 교통사고의 공간적 밀도 추정에 대해 네트워크 KDE(Kennel Density Estimation) 접근법을 제시
- Prasannakumar, V. 외 3명(2011): Moran의 공간적 자기상관법, Getis-Ord Gi* 통계 및 포인트 커널 밀도 함수를 사용하여 사고 및 핫스팟 공간 밀도의 클러스터링 수행, 교통사고가 밀집되는 구역을 핫스팟 커널로 시각화
- Anderson, T. K. (2009): 교통사고 지역을 셀(cell) 밀도에 따라 구분한 후 K-평균 클러스터링을 사용하여 커널밀도 분석 결과를 핫스팟으로 시각화
- 손세린. 강영옥 (2017): 교통사고의 시공간특성 분석을 위해 여성운전자의 교통사고 특성을 스페이스 타임 큐브 모형에 넣어 분석한 연구가 최근 진행됨



교통사고의 시공간적 특성과 교통사고 영향요인을 통합적 관점에서 분석한 연구는 부재

1. 연구배경 및 목적

연구목적

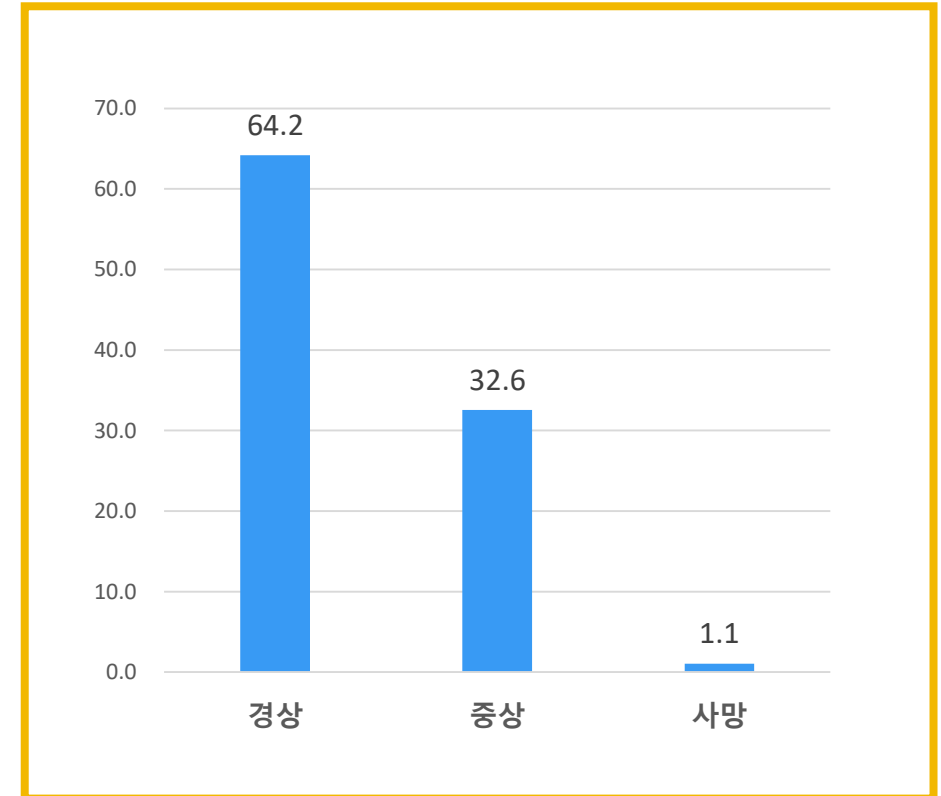
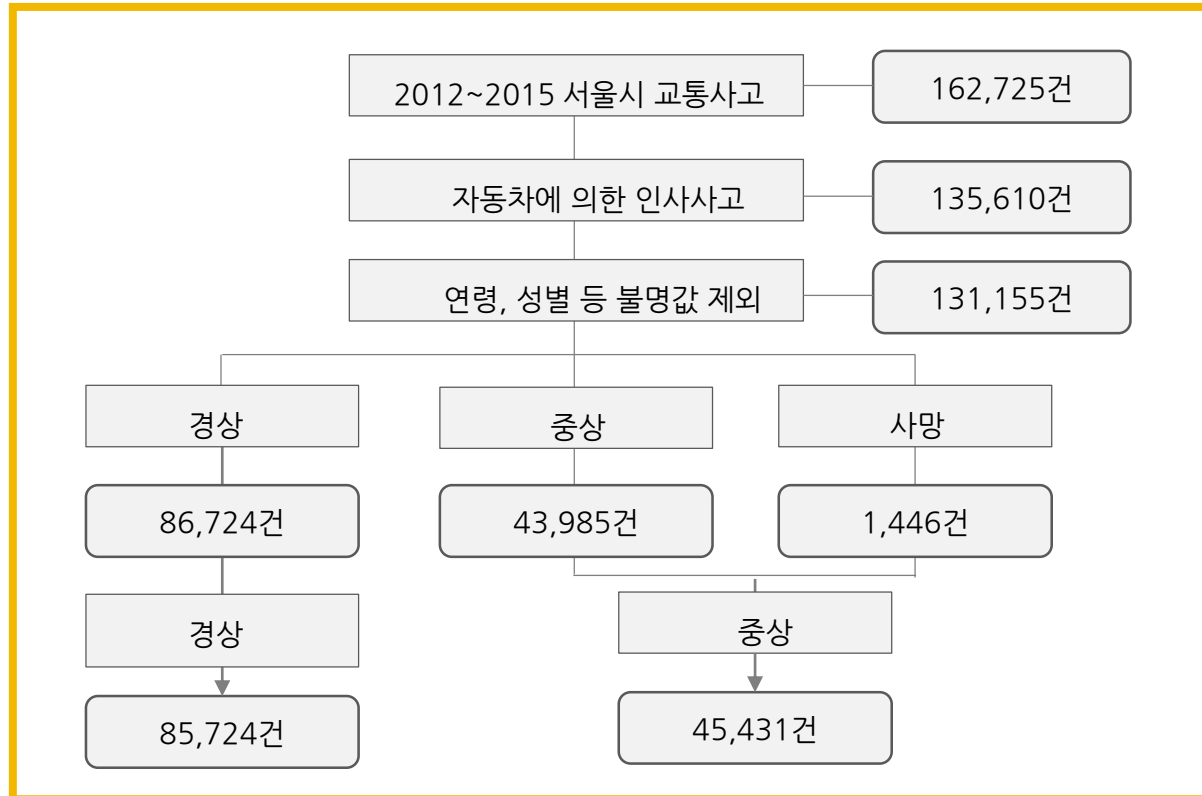
- ▶ 교통사고 심각도에 따른 시공간 특성 분석
- ▶ 교통사고 심각도에 영향을 미치는 주요요인 분석
- ▶ 심각한 교통사고를 유발하게 하는 규칙의 발견

2 연구대상, 범위, 방법

연구대상및범위

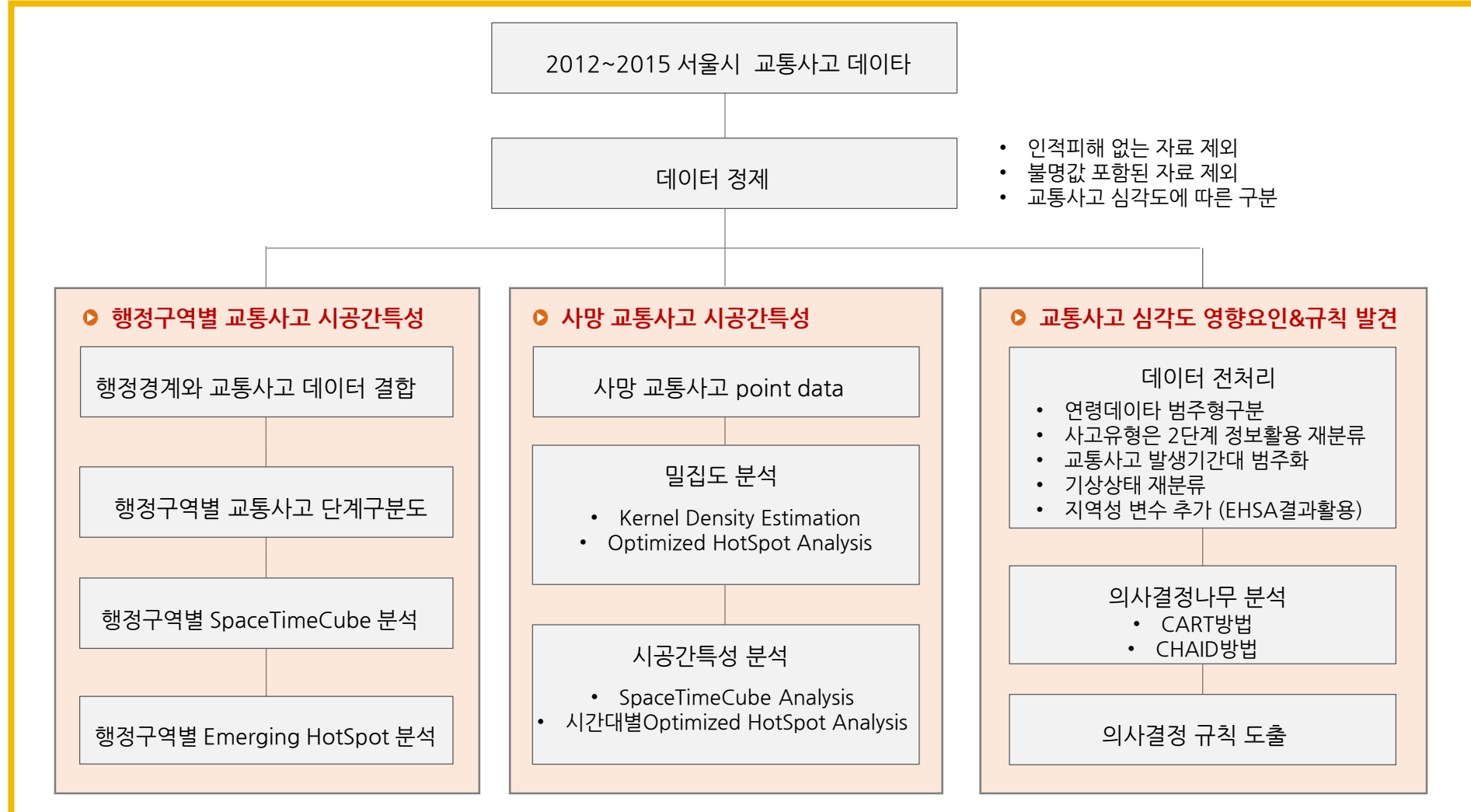
▶ 연구대상 및 범위 : 2012~2015년 사이 서울시에서 발생한 교통사고

- 행정구역경계_수도권 (통계청, 2015)
- 교통사고조사자료_수도권 (경찰청, 2012~2015)
- 사망교통사고정보_전국 (경찰청, 2012~2015)



2 연구대상,범위,방법

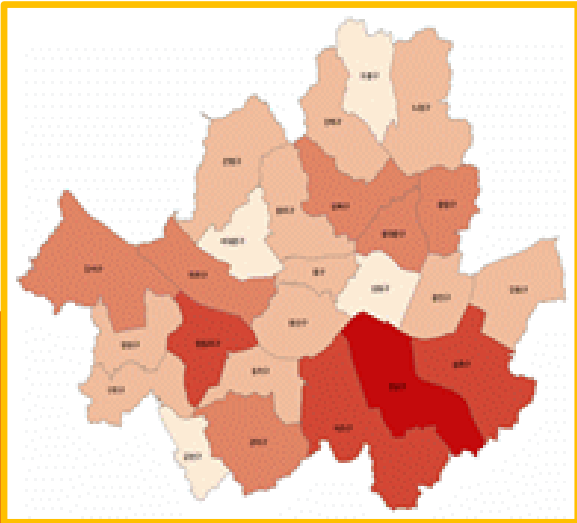
연구방법



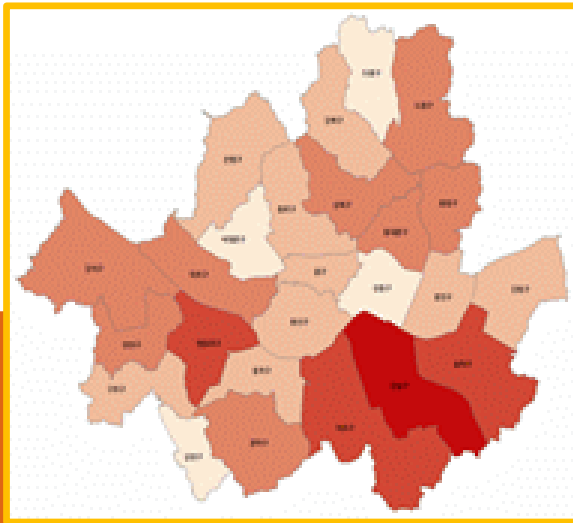
3. 주요연구결과

서울시동별 교통사고 집중지역

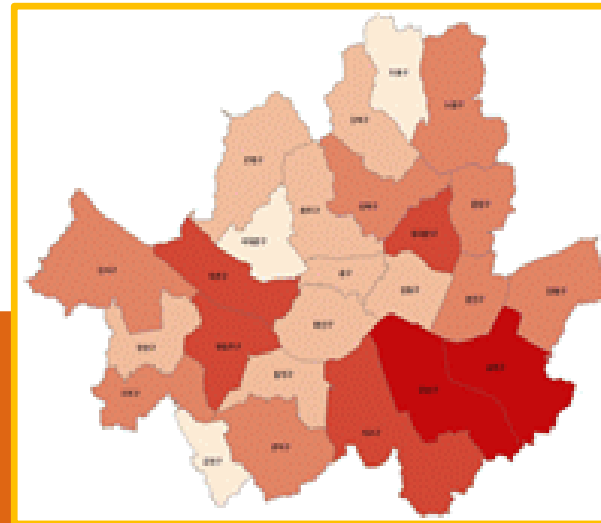
- ▶ 2012 ~ 2015년 사이 서울시 교통사고는 강남구에서 발생건수가 가장 많으며, 서초구, 송파구, 영등포구 순으로 많음
- ▶ 경상사고의 경우 강남3구와 영등포구에 많음
- ▶ 중상사고의 경우 경상사고와 유사하나 분포지역이 넓음
- ▶ 사망사고의 경우 영등포구가 가장 많으며, 도심보다 외곽에서 많으며, 한강이남의 상업활동이 많은 곳에 집중



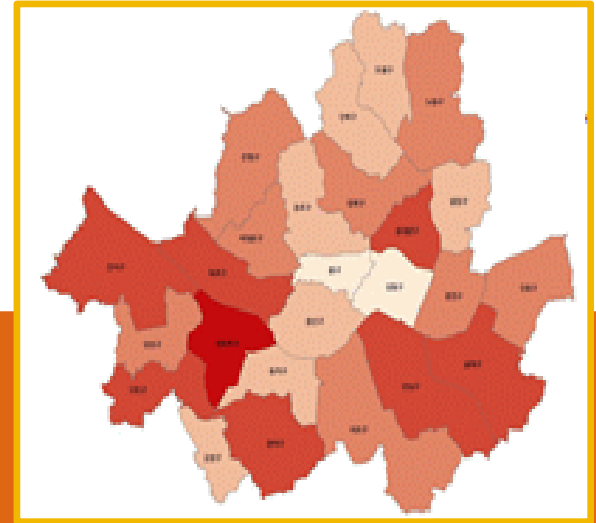
전체 교통사고 건수



경상 교통사고 건수



중상 교통사고 건수

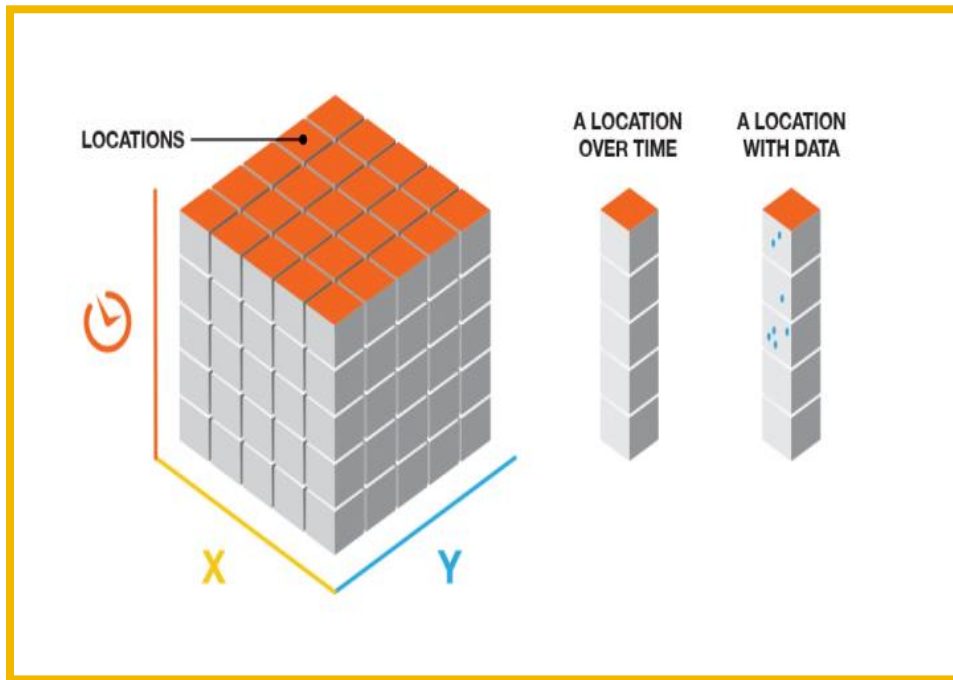


사망 교통사고 건수

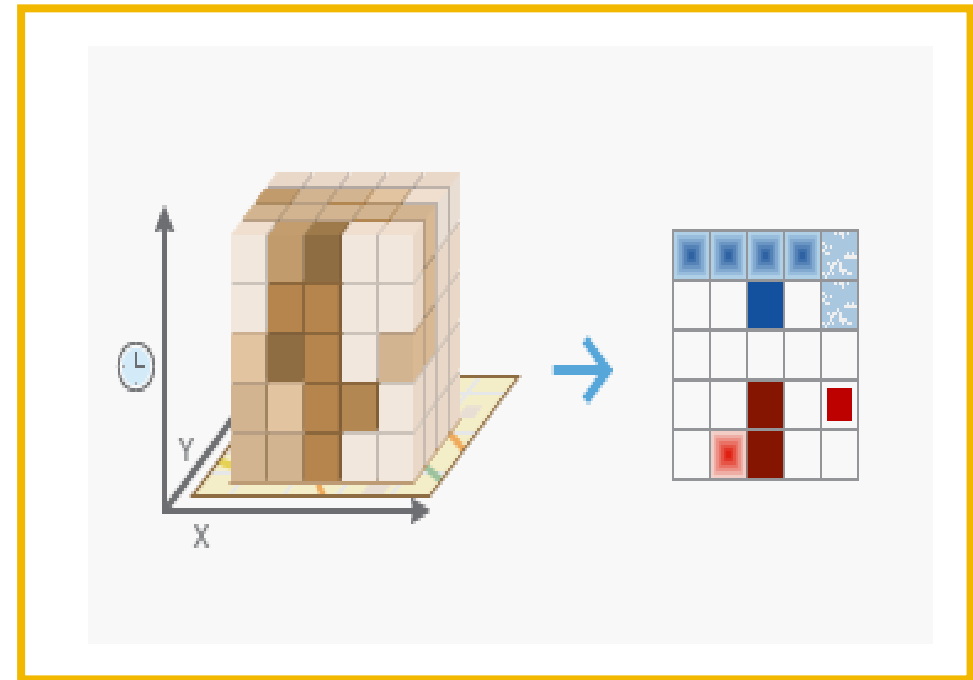
3. 주요연구결과

서울시동별 교통사고 집중지역 : SpaceTimeCube분석의 개념

- ▶ 사건발생 지점과 시간정보를 포함하는 데이터를 설정한 시간간격의 스페이스 타임 빈에 포함시켜 3차원 시공간 특성 분석
- ▶ 발생핫스팟을 통해 특정지역의 점밀도가 시간흐름에 따라 어떻게 변화하는지 경향성 분석



스페이스 타임 큐브 (SpaceTime Cube)



발생 핫스팟 (Emerging HotSpot Analysis)

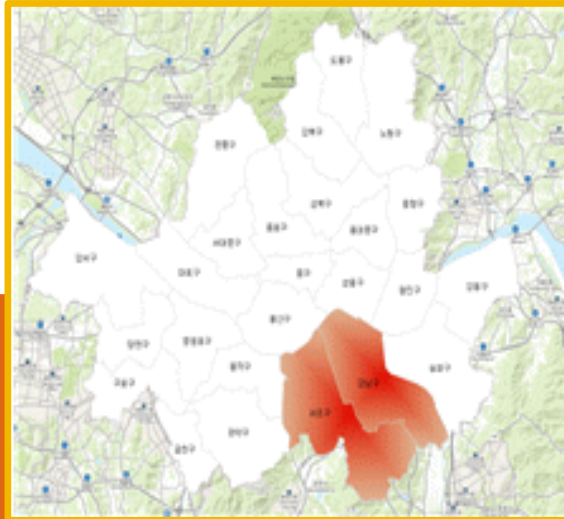
3. 주요연구결과

서울시동별 교통사고 집중지역:시공간패턴

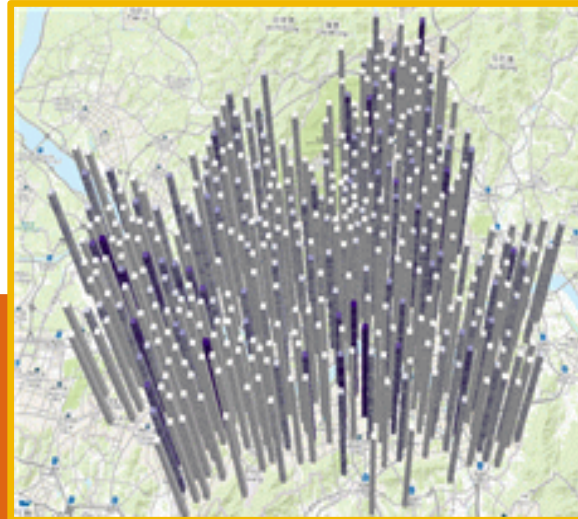
- ▶ 강남구와 서초구는 교통사고가 심화되는 ‘심화되는 핫스팟(Intensifying Hotspot)’ 지역으로 나타남
- ▶ 지속적인 핫스팟 (■: Persistent Hot Spot): 논현2동과 삼성2동
- ▶ 심화되는 핫스팟 (■: Intensifying Hot Spot): 반포동, 서초동, 역삼동, 도곡1동
- ▶ 산발적인 핫스팟 (■: Sporadic Hot Spot): 논현1동, 도곡2동, 신도림동과 망우동
- ▶ 산발적인 콜드 스팟(■: Sporadic Cold Spot): 관악구 삼성동은 유일하게 산발적으로 교통사고가 타지역에 비해 적게 나타남



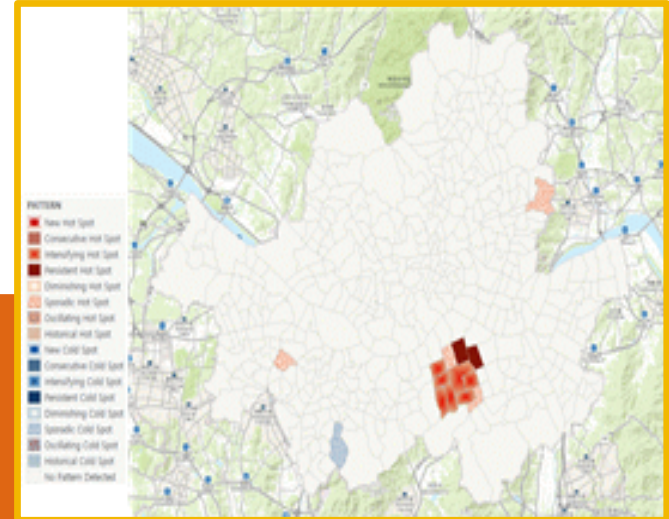
구별 스페이스타임큐브



구별 발생 핫스팟



동별 스페이스타임큐브

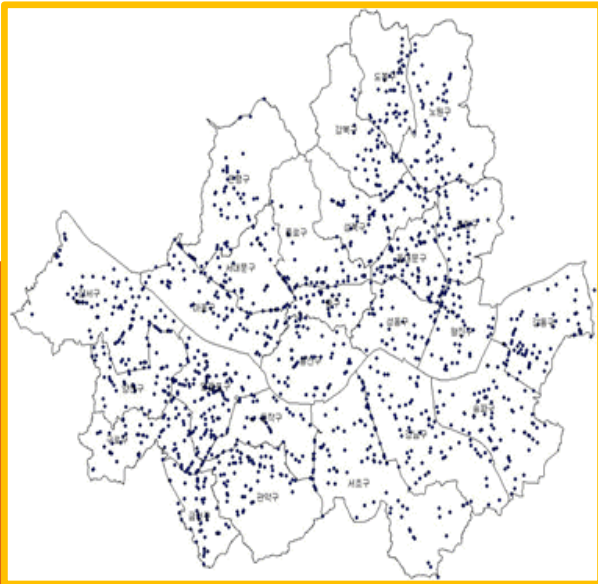


동별 발생 핫스팟

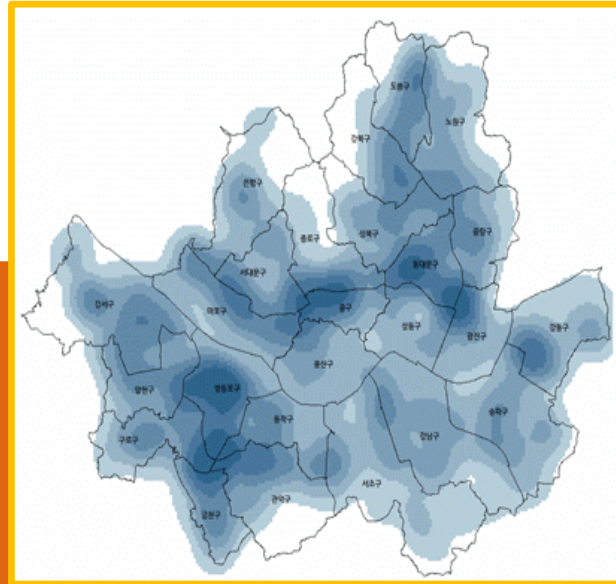
3. 주요연구결과

서울시사망교통사고 집중지역:공간적 특성

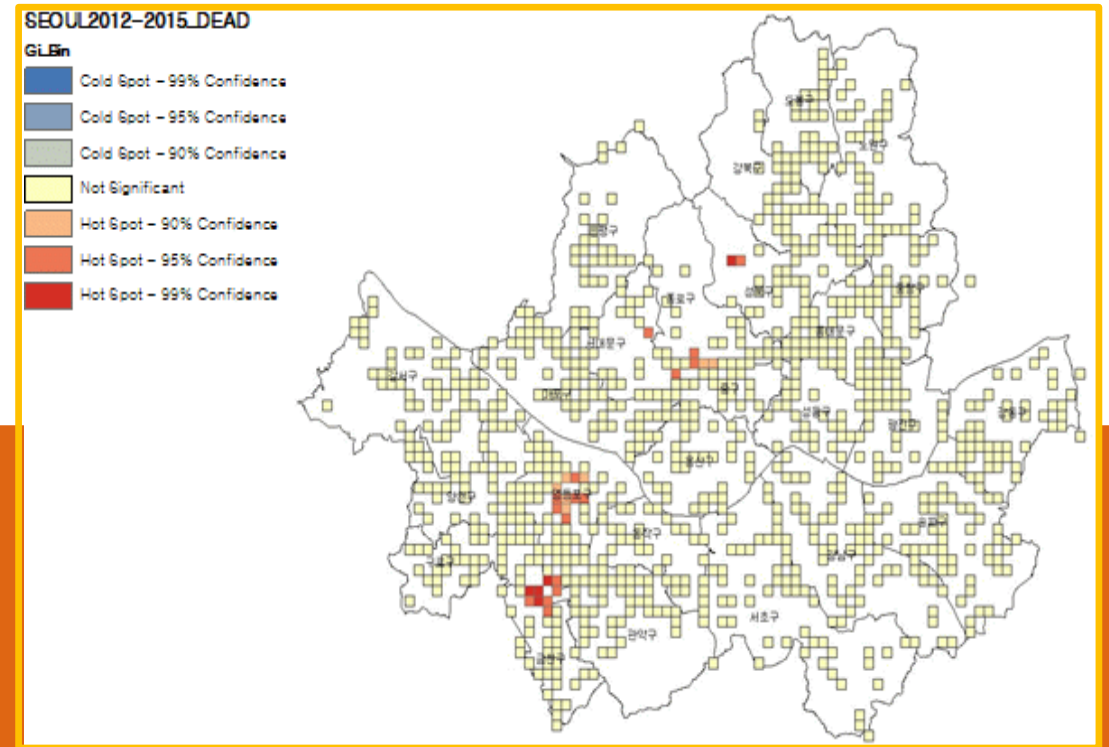
▶ 영등포구, 구로구, 종로구, 중구, 성북구 일부지역에 통계적으로 유의미한 사망 교통사고 밀집지역이 나타남



사망 교통사고 발생지점



사망 교통사고 커널분석

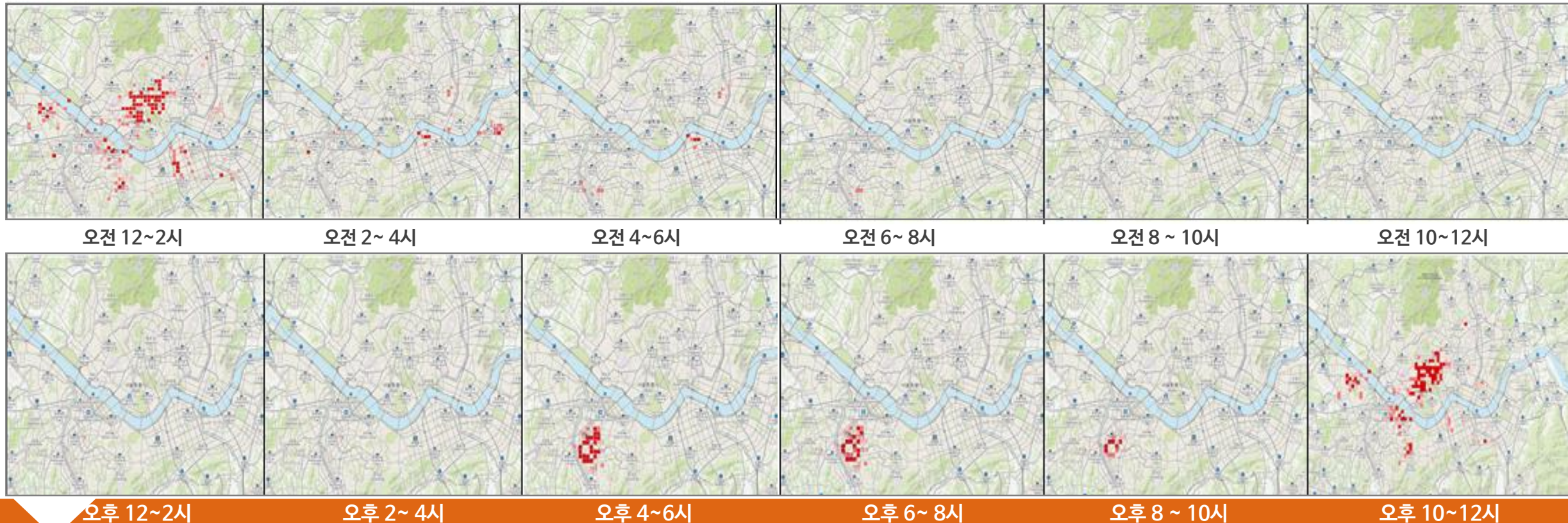


사망 교통사고 핫스팟분석

3. 주요연구결과

서울시사망교통사고 집중지역:시공간적 특성

- ▶ 자정을 넘긴 밤 12시에서 새벽 2시 사이에 중구, 종로구, 강남구의 유흥과 상업시설 집중지역, 강서구의 김포공항 IC부근과 인천국제 공항 고속도로 주변, 금천구의 서해안 고속도로 주변을 따라 가장 뚜렷한 핫스팟이 나타남
- ▶ 오전과 낮시간대에는 핫스팟이 없다가 오후 4시부터 8시 사이에 구로구의 디지털단지 주변의 가산로 부근에서 핫스팟 나타남
- ▶ 밤 10시 이후 종로, 중구, 마포구의 유흥지역, 강서구 일부지역, 영등포구과 관악구 서울대역 부근 일대에 핫스팟이 나타남



3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 의사결정나무분석의 개요

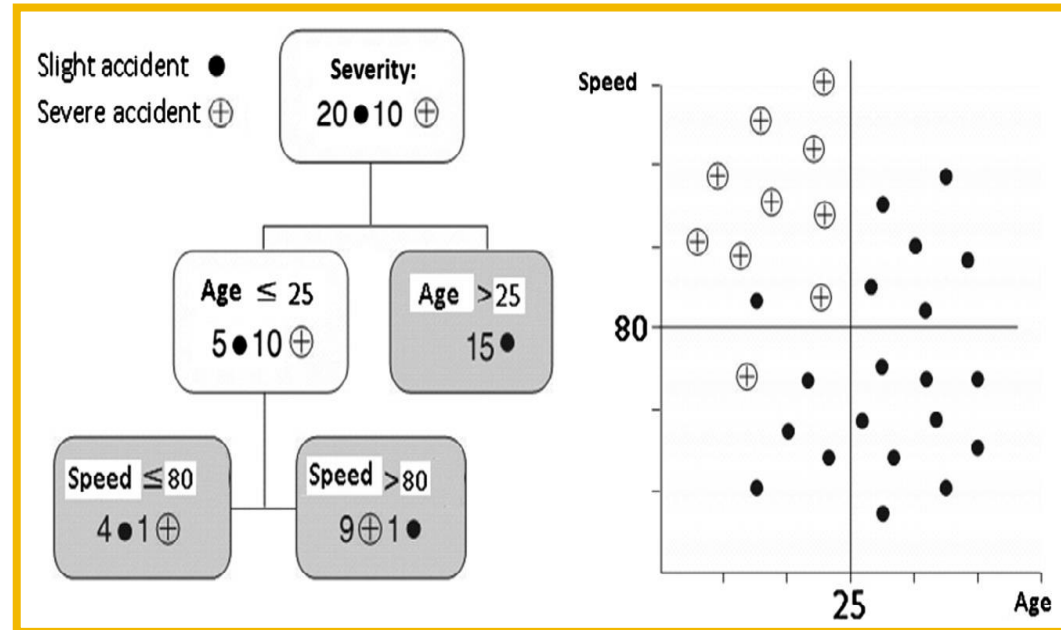
▶ **교통사고 데이터를 이용한 의사결정나무 생성** : 동질적인 목표 변수 그룹을 만들기 위해 데이터 분류

- 종속변수 : 교통사고 심각도, 독립변수 : 연령과 속도

▶ **의사결정 규칙의 생성**

- 만약(if)에 어떠한 상황에서는 어떤 결과를 낳게 된다는 규칙(rule)을 생성

- 만약(if) 나이가 25세 미만($age \leq 25$)이고, 차량속도가 80Km/h이하($speed \leq 80$)이면 교통사고는 가벼운 사고일 가능성이 높다



자료출처 : Abellan, J. et.al, (2013)

의사결정나무 구조와 분류 예시

3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙발견 : 의사결정나무분석알고리즘

▶ 의사결정나무 분석 알고리즘

- 분리기준, 정지규칙, 가지치기에 서로 다른 형성기준이 있고, 이의 결합에 따라 서로 다른 알고리즘을 생성

▶ 의사결정나무 분석 알고리즘 비교

	CHAID	CART	QUEST
목표변수	질적변수, 양적변수	질적변수, 양적변수	명목형질적변수
예측변수	질적변수, 양적변수	질적변수, 양적변수	질적변수, 양적변수
분리기준	F검정, 카이제곱통계량	지니계수감소	F검정, 카이제곱통계량
분리개수	다지분리	이지분리	이지분리

자료출처: 강병서·김계수, 2013

3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 데이터 전처리 및 변수 설정

▶ 교통사고 데이터의 분류

구분	변수	형태	속성값
인적요인	연령	연속형	
	성별	범주형	남, 여, 불명
	음주운전	범주형	정상운전, 음주운전, 불명
	사고유형	범주형	차대차, 차대사람, 차량단독
	법규위반	범주형	운전자법규위반, 보행자 과실, 정비불량
자동차요인	차량종별	범주형	승용차, 승합차, 이륜차, 자전거, 화물차, 기타
도로 및 환경적 요인	계절	범주형	봄, 여름, 가을, 겨울
	시간	연속형	
	요일	범주형	주중, 주말
	기상상태	범주형	맑음, 흐림, 안개, 비, 눈, 기타
	도로형태	범주형	교차로, 단일로, 고가도로 위, 기타
	도로종류	범주형	고속국도, 특별광역시도, 기타
	지역특성	범주형	교통사고 심화지역, 일반지역

3. 주요연구결과

교통사고 심각도 영향요인 & 규칙발견 :데이터 전처리 및 변수 설정

▶ 교통사고 데이터 정제 : 변수 설명표

번호	변수	형태	분류기준	사고심각도		
				빈도 (건)	경상(%)	중상(%)
1	연령	범주형	1. 20세미만 2. 20~65세 3. 65세미만	4,767 114,577 11,811	63.8 65.6 64.1	36.2 34.4 35.9
2	성별	범주형	1. 남 2. 여	112,170 18,985	65.3 65.5	34.7 34.5
3	음주운전	범주형	1. 음주운전 2. 정상운전 3. 측정불능	12,431 118,672 52	70.8 64.8 71.2	29.2 35.2 28.8
4	사고유형 (sago1)	범주형	1. 차대차 2. 차대사람 3. 차량단독	91,139 38,587 1,429	72.2 49.8 47.4	27.8 50.2 52.6
5	법규위반	범주형	1. 중대법규위반 2. 일반법규위반 3. 안전운전불이행 4. 기타	35,074 19,959 71,523 4,599	59.5 77.7 65.0 61.9	40.5 22.3 35.0 38.1
6	차량종별	범주형	1. 승용차 2. 승합차 3. 이륜차 4. 자전거 5. 화물차 6. 불명	92,909 8,533 8,956 4,997 10,833 4,927	67.9 58.7 59.2 59.7 58.9 59.5	32.1 41.3 40.8 40.3 41.1 40.5
7	계절	범주형	1. 봄 2. 여름 3. 가을 4. 겨울	33,668 33,736 34,604 29,147	65.0 65.4 64.5 66.7	35.0 34.6 35.5 33.3

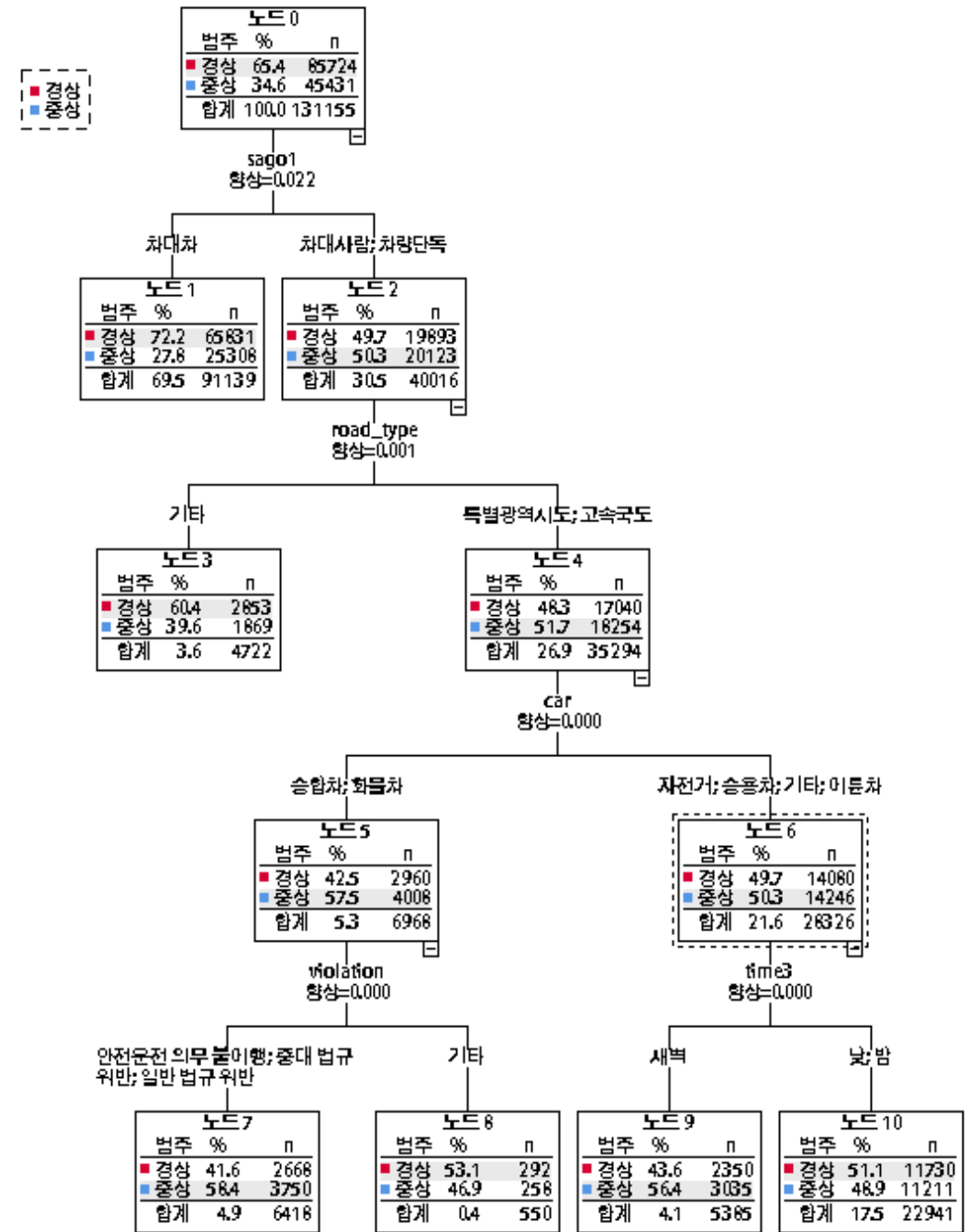
번호	변수	형태	분류기준	사고심각도		
				빈도 (건)	경상(%)	중상(%)
8	시간	범주형	1. 새벽 (0:00~6:00) 2. 낮 (6:00~18:00) 3. 밤 (18:00~24:00)	23,359 73,463 34,333	65.1 64.9 66.6	34.9 35.1 33.4
9	주야	범주형	1. 주 2. 야	66,101 65,054	64.8 66.0	35.2 34.0
10	요일	범주형	1. 주중 2. 주말	96,176 34,979	65.0 66.3	35.0 33.7
11	기상상태	범주형	1. 맑음 2. 흐림 3. 악천후(안개, 비, 눈) 4. 기타	112,428 6,635 10,687 1,405	65.5 63.7 65.2 67.0	34.5 36.3 34.8 33.0
12	도로형태	범주형	1. 교차로 2. 단일로 3. 기타	51,082 75,847 4,226	66.0 64.9 66.6	34.0 35.1 33.4
13	도로종류	범주형	1. 고속국도 2. 특별광역시도 3. 기타	303 122,853 7,999	64.0 65.4 65.0	36.0 34.6 35.0
14	지역성	범주형	1. 사고심화지역 2. 일반지역	18,915 112,240	68.9 64.8	31.1 35.2
	계			100.0 (131,155)	65.4 (85,724)	34.6 (45,431)

3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과 의사결정 규칙

▶ CART방법에 의한 의사결정나무

- 이진분류, 지니지수를 분리기준으로 사용
- 과적합 방지를 위해 부모노드 최소건수 1,000개, 자식노드 최소건수 500개
- 11개 노드, 6개의 터미널 노드, 4개의 계층 생성
- 분류정확도 66.7%

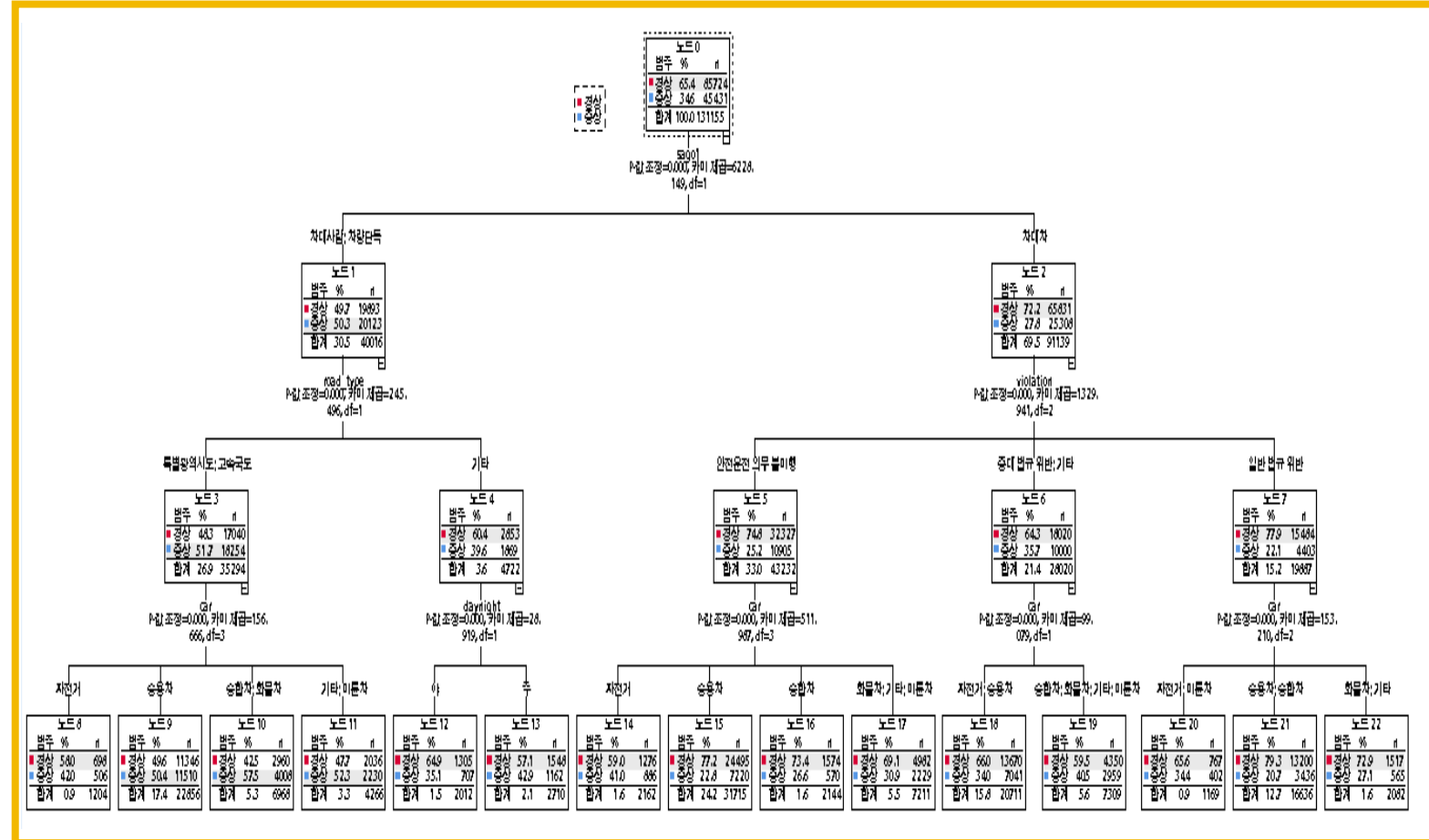


3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과 의사결정 규칙

▶ CHAID방법에 의한 의사결정나무

- 카이제곱검정통계량, 다지분류
- 과적합 방지를 위해 부모노드 최소건수 1,000개, 자식노드 최소건수 500개
- 23개 노드, 15개의 터미널 노드, 3개의 계층 생성
- 분류정확도 66.4%



3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과 의사결정 규칙

▶ 분류정확도

- 분류정확도는 CART알고리즘 적용시 66.7 %, CHAID알고리즘 적용시 66.4%로 CART알고리즘 적용시 분류정확도는 약간 높게 나타남
- 그러나 중상사고에 대한 분류 정확도는 CHAID알고리즘이 높게 나타남

	예측		
관측	경상	중상	정확도(%)
경상	80,706	5,018	94.1%
중상	38,646	6,785	14.9%
전체(%)	91.0%	9.0%	66.7%

CART 알고리즘

	예측		
관측	경상	중상	정확도(%)
경상	69,382	16,342	80.9%
중상	27,683	17,748	39.1%
전체(%)	74.0%	26.0%	66.4%

CHAID 알고리즘

3. 주요연구결과

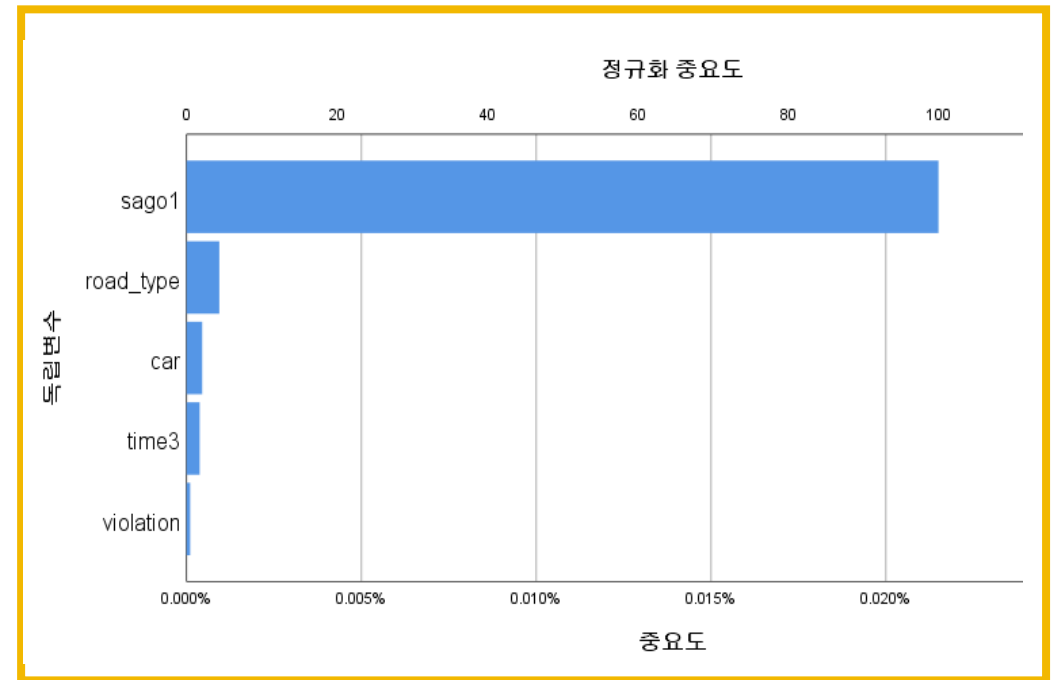
교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과의사결정규칙

▶ 교통사고 심각도에 영향을 미치는 요인

- 교통사고 심각도를 설명하는데 가장 중요한 독립변수는 사고유형으로 나타났으며, 도로종류, 차량종별, 교통사고 발생시간, 법규위반 등이 주요
- CART방법과 CHAID방법이 동일하게 나타남

독립변수	중요도	정규화중요도
사고유형(sago1)	.022	100.0%
도로종류(road_type)	.001	4.4%
차량종별(car)	.000	2.1%
시간(time3)	.000	1.7%
법규위반(violation)	.000	0.5%

독립변수 중요도



독립변수 정규화 중요도

3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과 의사결정 규칙

▶ 의사결정 규칙의 도출을 위한 기준

- 교통사고 심각성과 관련된 의미 있는 의사결정 규칙을 도출하기 위해 만약 A라면 B 라는 원칙 적용을 위한 파라미터의 최소 임계치와 원칙이 적용되는 확률 값의 임계치를 정함
- S (support; 지지도) : 데이터 셋에서 “A 와 B”가 함께 나타나는 비율이며, 최소값은 0.6%이상
- Po (Population) : 데이터 셋에서 “A”가 나타나는 비율이며, 최소값은 1% 이상
- P (Probability) : 규칙이 정확할 확률이며 P값은 ‘S/Po’로 산출되는데 최소값은 50%이상을 선정

▶ 교통사고 심각도에 영향을 미치는 요인에 대한 규칙

	규칙 (If)	결과 (Then)	S%	Po%	P%
1	사고유형=차대사람or차량단독;도로종류=고속국도or특별광역시도;차량종별=승합차or화물차;법규위반=안전운전의무불이행or중대법규위반or일반법규위반	중상	2.9	4.9	58.4
2	사고유형=차대사람or차량단독;도로종류=고속국도or특별광역시도;차량종별=승합차or화물차	중상	3.1	5.3	57.5
3	사고유형=차대사람or차량단독;도로종류=고속국도or특별광역시도;차량종별=자전거or승용차or이륜차or기타;시간=새벽	중상	2.3	4.1	56.4
4	사고유형=차대사람or차량단독;도로종류=고속국도or특별광역시도;차량종별=이륜차or기타	중상	1.7	3.3	52.3
5	사고유형=차대사람or차량단독;도로종류=고속국도or특별광역시도;차량종별=승용차	중상	8.8	17.4	50.4

* 규칙 1, 2, 3은 CRT방법, 규칙 2, 4, 5는 CHAID방법에 따라 산출됨.

3. 주요연구결과

교통사고 심각도 영향 요인 & 규칙 발견 : 모형구축과 의사결정 규칙

▶ 교통사고 심각도에 영향을 미치는 요인에 대한 규칙

1번 규칙:

- 사고유형이 차대 사람이거나 차량단독이면서 도로종류가 고속국도나 특별광역시도이고, 차량종류가 승합차나 화물차이면서 법규위반의 유형이 안전운전 불이행이거나, 중대법규위반, 일반법규위반이 아니면 교통사고 심각도가 가장 높음.
- 이 규칙에서 법규위반에 대한 규칙은 법규위반의 종류가 기타로 분류되는 값을 분기하기 때문에 실제로는 2번 규칙과 유사

2번 규칙:

- 사고유형이 차대 사람이거나 차량단독이면서 도로종류가 고속국도나 특별광역시도이고, 차량종류가 승합차나 화물차인 경우 교통사고는 중상으로 분류
- **고속도로나 특별광역시도와 같이 폭원이 넓고 차량속도가 높은 곳에서 승합차나 화물차에서 중상 이상의 교통사고가 일어날 가능성이 높음**

3번 규칙:

- 사고유형이 차대 사람이거나 차량단독이면서 도로종류가 고속국도나 특별광역시도이고, 차량종류가 승용차, 자전거, 이륜차, 기타 등이면 새벽에 교통사고 심각도가 높은 사고가 일어나는 것으로 분류
- **고속도로나 광역시도와 같이 폭원이 넓고 차량속도가 높은 곳에서 승용차, 자전거, 이륜차 등의 경우 새벽시간에 심각한 교통사고 발생 가능성이 높은 것으로 분류**

4번 규칙:

- 사고유형이 차대 사람이거나 차량단독이면서 도로종류가 고속국도나 특별광역시도이고, 차량종류가 이륜차이거나 기타인 경우 교통사고는 중상으로 분류
- 중상으로 분류하는 확률 값은 52.3%로 그다지 높지 않음

5번 규칙:

- 사고유형이 차대 사람이거나 차량단독이면서 도로종류가 고속국도나 특별광역시도이고, 차량종류가 승용차인 경우 교통사고는 중상으로 분류
- 이러한 경우가 나타나는 데이터는 전체의 17.4%이지만 이 경우에 중상으로 분류하는 확률은 50.4%로 확실적인 유의미도는 떨어짐
- 4번과 5번 규칙은 사고유형이 차대 사람이거나 차량단독의 경우 화물차나 승합차의 경우 심각한 교통사고가 될 확률이 가장 높게 나타나지만 자전거를 제외하면 이륜차나 기타의 경우, 혹은 승용차의 경우도 심각한 교통사고가 될 확률이 높음을 나타냄

4. 결론 및 정책제언

연구결과 요약

▶ 동별 교통사고 집중지역 : 공간적 특성

- 2012년에서 2015년까지 지난 4년간 서울시 교통사고는 강남구에서 발생건수가 가장 많으며, 그 다음이 서초구, 송파구, 영등포구 등에서 교통사고가 많이 발생
- 성동구, 서대문구, 도봉구에서는 상대적으로 교통사고 발생건수가 작았음.
- 교통사고 심각도에 따라 구분해 보면 경상사고의 경우 강남 3구와 영등포구에서 많이 발생
- 중상사고의 경우 경상사고와 다르지 않지만 보다 분포지역이 널리 퍼짐
- 사망사고의 경우 영등포구를 중심으로 널리 퍼지게 분포됨.
- 도심부 보다는 외곽지역에서 많이 발생하며 특히 한강 이남의 상업 활동이 많은 곳에서 집중됨

▶ 동별 교통사고 집중지역 : 시공간적 특성

- 교통사고를 월단위의 시간흐름에 따라 분석한 결과 강남구와 서초구는 ‘심화되는 핫스팟(Intensifying Hotspot)’ 지역으로 나타남
- 지속적인 핫스팟 (Persistent Hot Spot) : 논현2동과 삼성2동
- 심화되는 핫스팟(Intensifying Hot Spot) : 반포동, 서초동, 역삼동, 도곡1동
- 산발적인 핫스팟 (Sporadic Hot Spot) : 논현1동, 도곡2동, 신도림동과 망우동
- 산발적인 콜드 스팟(Sporadic Cold Spot) : 관악구 삼성동은 유일하게 산발적으로 교통사고가 타지역에 비해 적게 발생
- 동별 핫스팟 분석결과는 구별 핫스팟 분석과 크게 다르지 않았지만 서초구와 강남구내에서 특히 논현동, 서초동, 반포동, 역삼동, 도곡동, 삼성동 등 특정지역에 교통사고가 집중되어 있고 시간이 흐르면서 교통사고가 감소하지 않으며 지속적으로 발생이 심화되어 더욱 주의를 요하는 지역임을 알 수 있었음.
- 구로구 신도림동이나 도봉구 망우동처럼 산발적인 핫스팟이 나타나 구별 분석에서 나타나지 않지만 주의를 요하는 지역도 확인할 수 있었음

4. 결론 및 정책제언

연구결과 요약

▶ 사망 교통사고 집중지역 : 시공간적 특성

- 사망 교통사고는 영등포구, 구로구, 중구, 동대문구, 광진구, 강동구 일대에서 집중되어 나타남
- 통계적으로 유의미한 사망사고 집중지역은 영등포구, 구로구, 종로구, 중구, 성북구 일부지역에 작게 나타남
- 시간대별로는 밤 12시 ~ 새벽 2시 : 중구, 종로구, 강남구의 유흥과 상업시설 집중지역, 강서구의 김포공항 IC부근과 인천국제공항고속도로 주변, 금천구의 서해안 고속도로 주변을 따라 핫스팟
- 새벽 2시 ~ 4시 : 강남구 신사동, 압구정동 일대, 강동구 천호대로와 강동대로 주변지역
- 새벽 4시 ~ 6시 : 강남구 신사동, 압구정동 일대, 구로구 일대에서 작은 규모의 약한 핫스팟이 나타남
- 오전 6시 이후 ~ 낮 시간 : 핫스팟이 거의 도출되지 않음
- 퇴근 시간대인 오후 4시 ~ 8시 : 구로구의 디지털단지 주변의 가산로 부근에서 뚜렷한 핫스팟이 나타남. 이후 10시 까지 핫스팟이 약해짐
- 밤10시 ~ 12시 : 종로, 중구, 마포구의 유흥지역, 강서구 일부 지역, 영등포과 관악구 서울대역 부근 일대로 핫스팟이 넓게 퍼져가며 나감
- 시간대별 사망사고 집중지역의 변화는 공간적 집중만을 분석했을 때는 구체적으로 나타나는 않는 현상이었음

▶ 교통사고 심각도에 영향을 미치는 요인과 규칙

- 교통사고 심각도에 영향을 미치는 독립변수 : 차대차, 차대사람, 차량단독을 나타내는 사고유형이 가장 중요한 역할을 하며 도로의 종류, 차량의 종류, 교통사고 발생 시간, 법규위반 종류 등의 순으로 중요도가 나타남
- 심각한 교통사고가 발생하는 몇 가지 규칙 :
차대사람이나 차량단독으로 사고가 나는 경우 고속도로나 특별광역시도와 같이 폭원이 넓고 차량속도가 높은 곳에서 승합차나 화물차에서 중상의 교통사고가 일어날 가능성이 높으며, 동일한 상황에서 승합차나 화물차가 아닌 승용차, 자전거, 이륜차 등의 경우에는 새벽시간에 심각한 교통사고로 이어질 가능성이 높음

4. 결론 및 정책제언

정책제언

▶ 교통사고 데이터 제공단위 개선

: 교통사고 발생지점 정보를 현재의 동단위 데이터에서 사고지점별 데이터로 제공필요

: 동단위로 시공간 분석을 수행하는 경우 의미있는 분석 어려움

▶ 교통사고 발생 시 수집되는 데이터의 내용 (속성정보 or포함정보) 개편 필요

: 의사결정나무 분류정확도는 66.7%로 낮은편은 아님

: 중상교통사고의 분류정확도는 상대적으로 낮았음

: 교통사고 발생 상황에 대한 구체적 정보 필요 - 해당도로의 폭원, 차량속도, 시야, 신호등 여부, 안전시설 여부 등

▶ 교통사고 데이터가 지점별로 제공되면 다른 정보와 연계하여 상황정보 구축가능

: 발생지점의 도로여건, 해당도로의 속도정보, 발생지점의 토지이용 등과 연계하여 상황정보 구축가능하며, 구체적 분석이 가능할 것으로 판단됨

▶ 데이터 개방과 관련된 제도개선이 지속되길 바램

참고문헌(1)

- 강병서·김계수, 2013, 「사회과학 통계분석」, 한나래 아카데미.
- 남기성·안병준·김주환, 2004, “데이터마이닝을 이용한 복합적 교통사고 요인 추출”, 『Journal of the Korean Data Analysis Society』, 6(2): 519-525.
- 박용우·백세흠·박신행·권오훈, 2016, “의사결정나무를 이용한 고속도로 공간구간 사고 심각도에 관한 연구”, 『대한교통학회지』, 34(6): 535~547.
- 박재영·김도경, 2011, “사회경제지표를 반영한 거시적 교통사고 분석모형 개발”, 『한국교통연구원 교통연구』, 18(1):53~62.
- 박준태·김용석·이수범, 2009, “로짓모형을 이용한 고령운전자 고속도로 교통사고 특성 분석 연구”, 『한국도로학회논문집』, 11(4): 1~7.
- 박준태·이수범·이동민, 2011, “도시부 도로구간 사고예측모형 개발”, 『한국교통연구원 교통연구』(1):63~73.
- 박창이·김용대·김진석·송종우·최호식, 2013, 「R을 이용한 데이터 마이닝」, 교우사.
- 성병준·유환희, 2014, “진주시 교통사고 공간적 군집분석”, 『한국지형공간정보학회 학술대회』, 169-172.
- 손세린·강영옥, 2017, “서울시 여성운전자 교통사고의 시공간 특성분석”, 『한국지도학회지』, 17(2): 89-98.
- 이건학, 2004, “GIS와 공간 데이터마이닝을 이용한 교통사고의 공간적 패턴 분석 -서울시 강남구를 사례로”, 『대한지리학회지』, 39(3): 457-472.
- 이문영, 2013, “의사결정나무 모형을 활용한 교통사고 요인분석”, 명지대학교 교통공학과 박사학위논문
- 이재길, 2017, 「R프로그램에 기반한 다변량분석 및 데이터마이닝」, 황소걸음아카데미.
- 이재명·김태호·이용택·원제무, 2008, “CART분석을 이용한 교통사고예측모형의 개발”, 한국도로학회논문집, 10(1):31~39.
- 전우훈·조혜진, 2005, “야간 교통사고의 특성 분석 연구”, 『대한토목학회 학술대회』, (10): 4729-4732.
- 조나혜·강영옥, 2016, “로그데이터의 시공간 데이터마이닝 및 시각화 연구동향”, 『한국지도학회지』, 16(3): 15-27.
- 하오근·박동주·원제무·정철호, 2010, “고속도로 사고등급별 돌발상황 처리시간 예측모형 및 의사결정나무 개발”, 『한국ITS학회논문집』, 9(1): 101~110.
- 한상진·조운범·장수은, 2014, CART를 이용한 화물자동차 교통사고 특성분석, 『한국교통연구원 교통연구』, 21(4):87~103.
- Abdel Wahab, H. T. and Abdel-Aty, M. A., 2001, “Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections.”, *Transportation Research Record*, 17(46): 6-13.
- Abdel-Aty, M. A., and Radwan, A. E., 2000, “Modeling traffic accident occurrence and involvement”, *Accident Analysis & Prevention*, 32(5): 633-642.
- Abellan, J., Lopez, G. and de Ona, J., 2013, “Analysis of traffic accident severity using Decision Rules via Decision Trees”, *Expert Systems with Applications*, 40(15): 6047 ~ 6054.
- Anderson, T. K., 2009, Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3), 359-364.
- Bayam, E., Liebowitz, J., and Agresti, W., 2005, “Older drivers and accidents: A meta analysis and data mining application on traffic accident data”, *Expert Systems with Applications*, 29(3): 598-629.
- Chang, L. Y. and Chien, J. T., 2013, “Analysis of driver injury severity in truck involved accidents using a non-parametric classification tree model”. *Safety Science*, 51(1): 17-22.
- Chang, L. Y. and Wang, H. W., 2006, “Analysis of traffic injury severity: An application of non-parametric classification tree techniques.”, *Accident Analysis and Prevention*, 38(5), 1019-1027.
- De Ona, J., Lopez, G., Mujalli, R. O. and Calvo, F. J., 2013, “Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks.”, *Accident Analysis and Prevention*, 51: 1-10.
- DeJoy, D. M., 1992, “An examination of gender differences in traffic accident risk perception,” *Accident Analysis & Prevention*, 24(3), 237-246.

참고문헌 (2)

- Elvik, R., 2013, "Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies," *Accident Analysis & Prevention*, 60, 254-267.
- Graham, J., Irving, J., Tang, X., Sellers, S., Crisp, J., Horwitz, D., ... & Carey, D., 2015, "Increased traffic accident rates associated with shale gas drilling in Pennsylvania," *Accident Analysis & Prevention*, 74, 203-209.
- Kristensson, P. O., Dahlback, N., Anundi, D., Bjornstad, M., Gillberg, H., Haraldsson, J., ... and Stahl, J., 2009, "An evaluation of space time cube representation of spatiotemporal patterns", *IEEE Transactions on Visualization and Computer Graphics*, 15(4): 696-702.
- Massie, D. L., Campbell, K. L., and Williams, A. F., 1995, "Traffic accident involvement rates by driver age and gender", *Accident Analysis & Prevention*, 27(1): 73-87.
- Montella, A., Aria, M., D'Ambrosio, A. and Mauriello, F., 2012, "Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery.", *Accident Analysis and Prevention*, 49: 58-72.
- Montella, A., Aria, M., D'Ambrosio, A., and Mauriello, F., 2011, "Data mining techniques for exploratory analysis of pedestrian crashes.", *Transportation Research Record*, 2237: 107-116.
- Mujalli, R. O. and de Ona, J., 2011, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks.", *Journal of Safety Research*, 42(5), 317-326.
- Parker, D., West, R., Stradling, S., and Manstead, A. S., 1995, "Behavioural characteristics and involvement in different types of traffic accident", *Accident Analysis & Prevention*, 27(4): 571-581.
- Prasannakumar, V., Vijith, H., Charutha, R., and Geetha, N., 2011, "Spatio-temporal clustering of road accidents: GIS based analysis and assessment", *Procedia-Social and Behavioral Sciences*, 21: 317-325.
- Xie, Z., and Yan, J., 2008, "Kernel density estimation of traffic accidents in a network space", *Computers, Environment and Urban Systems*, 32(5): 396-406.
- Zhang, G., Yau, K. K., Zhang, X., & Li, Y., 2016, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis & Prevention*, 87, 34-42.
- ESRI ArcGIS Pro, <http://pro.arcgis.com>.
- 교통사고조사자료_도로교통공단 통합DB처, 2014
- 교통사고조사자료_수도권, 경찰청, 2012-2015
- 교통사고통계 요약, 도로교통공단, 2016
- 사망교통사고정보_전국, 경찰청, 2012-2015
- 집계구별 인구통계, 통계청, 2015
- 행정구역경계_수도권, 통계청, 2015



강영옥
이화여자대학교사회과교육과 교수
ykang@ewha.ac.kr