



범죄예측에서의 데이터마이닝 적용 가능성 연구

- 절도범죄를 중심으로

A Study on the Applicability of Data Mining for Crime Prediction : Focusing on Burglary

저자 (Authors)	방승환, 김태훈, 조현보 Seung-Hwan Bang, Tae-Hun Kim, Hyun-Bo Cho
출처 (Source)	한국컴퓨터정보학회논문지 19(12) , 2014.12, 309-317 (9 pages) Journal of the Korea Society of Computer and Information 19(12) , 2014.12, 309-317 (9 pages)
발행처 (Publisher)	한국컴퓨터정보학회
URL	http://www.dbpia.co.kr/Article/NODE06531494
APA Style	방승환, 김태훈, 조현보 (2014). 범죄예측에서의 데이터마이닝 적용 가능성 연구. 한국컴퓨터정보학회논문지, 19(12), 309-317.
이용정보 (Accessed)	경성대학교 210.110.180.*** 2016/11/12 19:26 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

이 자료를 원저작자와의 협의 없이 무단게재 할 경우, 저작권법 및 관련법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

The copyright of all works provided by DBpia belongs to the original author(s). Nurimedia is not responsible for contents of each work. Nor does it guarantee the contents.

You might take civil and criminal liabilities according to copyright and other relevant laws if you publish the contents without consultation with the original author(s).

범죄예측에서의 데이터마이닝 적용 가능성 연구 : 절도범죄를 중심으로

방승환*, 김태훈*, 조현보*

A Study on the Applicability of Data Mining for Crime Prediction : Focusing on Burglary

Seung-Hwan Bang*, Tae-Hun Kim*, Hyun-Bo Cho*

요약

최근, 범죄가 증가함에 따라 범죄를 예측하고 예방하는 것은 사회의 중요한 이슈이며 정부 및 지자체는 다양한 방법론을 활용하여 범죄를 사전에 막고자 노력하고 있다. 데이터마이닝은 범죄예측 및 예방에 활용되는 대표적인 방법론이며, 범죄 패턴 분석, 범죄 발생 예측 등 다양한 분야에서 연구되고 있다. 그러나 데이터마이닝의 결과가 범죄학에서의 범죄 환경요소와 어떤 관련이 있는지 혹은, 사건해결에 어떤 도움을 줄 수 있는지에 대한 연구는 이루어지고 있지 않다. 따라서 본 논문에서는, 범죄학에서 범죄의 발생과 범죄 환경요소들의 상호 관련성을 보이고 범죄 발생과 관련된 환경요소와 데이터마이닝에 활용되는 변수 간의 관계를 정의하고자 하였다. 또한, 국내 보호관찰소에서 보관되고 있는 절도범죄 데이터를 사용하여 실제로 데이터마이닝의 결과가 범죄 환경요소와 어떤 관련이 있는지를 보이기 위해 군집분석을 적용하였다. 그 결과 각 군집별로 범죄가 발생하는 환경에 차이가 있었으며, 이를 활용하여 데이터마이닝이 범죄학관점에서 범죄 예측 및 예방 활용에 유의미함을 보였다.

▶ Keywords : 범죄예측, 데이터마이닝, 범죄학, 범죄환경

Abstract

Recently, crime prediction and prevention are the most important social issues, and global and local governments have tried to prevent crime using various methodologies. One of the methodologies, data mining can be applied at various crime fields such as crime pattern analysis, crime prediction, etc. However, there is few researches to find the relationships between the results of data mining and crime components in terms of criminology. In this study, we introduced environmental criminology, and identified

•제1저자 : 방승환 •교신저자 : 조현보

•투고일 : 2014. 8. 24, 심사일 : 2014. 9. 30, 게재확정일 : 2014. 10. 28.

* 포항공과대학교 산업경영공학과(Dept. of Industrial and Management Engineering, POSTECH)

※ 본 논문은 미래창조과학부 및 정보통신산업진흥원의 SW컴퓨팅산업원천기술개발사업(10047146, 범죄 발생환경, 행동패턴 및 심리정보등 융합정보 적용형 엔트로피 필터링 예측분석기반의 실시간 범죄 예측·예방 시스템 개발)의 일환으로 수행되었습니다.

relationships between environment factors related with crime and variables using at data mining. Then, using real burglary data occurred in South Korea, we applied clustering to show relations of results of data mining and crime environment factors. As a result, there were differences in the crime environment caused by each cluster. Finally, we showed the meaning of data mining use at crime prediction and prevention area in terms of criminology.

▶ Keywords : Crime prediction, Data mining, Criminology, Crime environment

I. 서 론

사회가 발전하고 과학기술의 개발에 힘입어 삶의 질이 증가하고 동시에 범죄로 인한 피해도 따라 증가하고 있다. 또한 특정한 지역으로 인구가 집중되어 도시가 급격히 변화하고 이로 인해 범죄에 대응하고 예방하는 것이 쉽지 않다. 이러한 변화에 대응하여 정부 및 지자체는 다양한 기법을 활용하여 범죄를 예측하고, 예측된 정보를 활용하여 치안인력을 효율적으로 운용하고 범죄 사건을 해결하는데 많은 노력을 기울이고 있다[1]. 한편 범죄 사건을 해결하기 위해 범죄학 초기에는 가해자를 대상으로 한 연구가 이루어졌다면, 현재에는 범죄와 관련된 요소(범행 장소, 주위 환경 등)들을 함께 고려하기 시작하였다. GIS 정보를 활용하여 범행의 시공간적 패턴에 중점을 둔 연구가 진행되고 있는 반면[2], 범죄와 관련된 데이터에서 기존에 알지 못했던 새로운 정보를 추출하는데 중점을 둔 연구가 있다[3][4]. 위의 연구에서 활용되는 데이터의 경우, 과거에는 수집될 수 있는 데이터가 적었던 반면, 최근에는 다양한 방면에서 데이터가 수집되고 있어 종류가 다양해지고 방대해지고 있다. 그러나 데이터가 불완전하고 잡음이 많아 분석 정확도가 떨어지고 그 효율성이 낮아 활용하는 것이 쉽지 않다[5][6]. 이에 따라 데이터 분석 정확도를 높임과 동시에 의미 있는 지식을 추출하여 활용도를 높이기 위해 데이터마이닝이 주목 받기 시작하였다.

데이터마이닝은 일반적으로 방대한 양의 데이터로부터 유용한 정보를 추출하는 과정을 의미한다. 이를 통해 기존에 모르던 새로운 정보를 발견할 수 있으며 사건 해결을 위한 의사결정에 도움을 줄 수 있다. 데이터마이닝에 사용될 수 있는 데이터의 종류는 일반적으로 시공간 요소, 범주 환경요소 및 가해자 프로파일로 구분되며[7], 어떤 데이터마이닝 기법을 적용하느냐에 따라 다양한 특성의 정보(범죄 지역 특징, 가해

자의 행동 정보 등)를 얻을 수 있다[8]. 그러나 데이터마이닝을 통해 얻은 정보가 범죄학(Criminology)관점의 범죄 구성 요소와 어떤 의미가 있는지에 대한 연구는 이루어지고 있지 않아 범죄 해결에 어떤 도움을 주는지 명확하지 않다.

따라서, 본 연구에는 범죄예측에 활용되고 있는 데이터마이닝이 실제로 범죄 해결에 도움을 줄 수 있는지 알아보고, 어떤 의미가 있는지 알아보고자 한다. 이를 위해, 범죄학에서의 범죄 발생을 위한 구성 요소와 데이터마이닝에 활용되는 변수를 비교하고자 한다. 또한, 대표적인 데이터마이닝 기법과 그 특성에 대해 알아보고, 국내 범죄 사건 기록에 데이터마이닝 기법을 적용하여 새롭게 얻은 결과가 어떤 의미가 있는지 분석하였다.

II. 범죄 발생 환경

범죄학에서는 범죄의 발생과 주위 환경요소들이 상호 관련되어 있다는 연구가 기존에 많이 진행되었다. 대표적인 이론으로 일상활동이론(Routine activity theory), 범죄패턴이론(Crime pattern theory), 의식이론(Awareness theory) 등이 있다[6].

1. 일상활동이론

일상활동이론은 범죄가 발생하기 위해 동기가 부여된 가해자(Motivated offender), 보호능력의 부재(Absence of capable guardian) 및 적절한 표적(Suitable target)이라는 조건이 적절한 시간과 공간에 주어져야 한다는 이론이다[9]. 이 이론은 그림 1과 같이 간단히 도식화하여 표현할 수 있다. 세 가지의 범죄유발요인이 동일한 시간과 동일한 공간에 어떻게 나타나는지에 집중하고 있으며 각 요인이 상호작용하여 범죄가 발생한다고 설명한다.



그림 1. 일상활동이론 구성요소

2. 범죄패턴이론

범죄패턴이론은 공간적인 환경요소를 마디(Nodes), 경로(Paths), 가장자리(Edges)로 구분하고 범죄가 각 환경요소에 의해 어떻게 영향을 받는지에 대한 이론이다[10]. 마디는 사람이 출발하고 이동하여 도착하는 장소를 의미하며 집, 학교, 백화점 등이 해당한다. 경로는 사람이 출근하는 것과 같이 평소에 이용하는 길을 의미하며, 가장자리는 사람이 거주 지역이나 평소에 특정한 목적을 위해 움직이는 범위의 끝을 의미한다. 마디, 경로 및 가장자리의 예시는 그림 2와 같다. 범죄패턴이론에 따르면 범죄는 무작위로 발생하는 것이 아니라, 공간적인 환경과 개인 및 이웃의 생활 패턴에 영향을 받아 발생한다.

3. 의식이론

의식이론은 범죄는 4개의 구성요소인 피해자(Victim), 가해자(Offender), 시공간(Geo-temporal) 및 규제(Legal)로 구성되어 있다는 이론이다[11]. 이론에 의하면 범죄는 각 구성요소에 영향을 받기 때문에 각 요소에 집중하면 가해자의 행동이나 범죄의 특징을 이해할 수 있다. 특히, 각각의 요소뿐만 아니라 여러 요소가 복합적으로 작용하기 때문에 모든 요소를 함께 고려해주어야 한다.

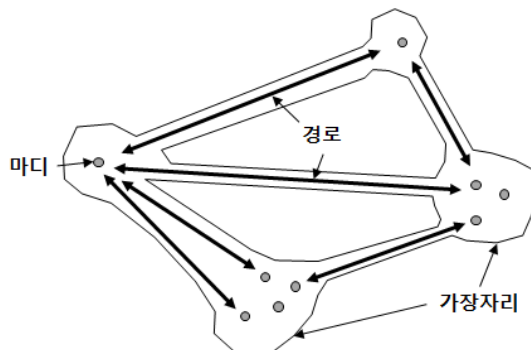


그림 2. 마디, 경로, 가장자리 예시

4. 범죄학에서의 데이터마이닝 활용

앞에서 제기한 범죄 이론에 의하면 범죄의 발생은 표 1과 같은 범죄 환경요소에 영향을 받는다. 일상활동이론 관점에서는 가해자와 피해자 그리고 주위 환경에 의한 보호능력의 부재가 갖추어 져야 범죄가 발생한다. 범죄패턴이론 관점에서는 공간적인 환경과 피해자 및 가해자의 패턴에 따라 범죄가 발생한다. 또한, 의식이론 관점에서는 피해자, 가해자 및 시공간 환경과 환경에 따른 규제로 범죄가 발생한다.

표 1. 범죄학에서의 범죄 환경요소

	범죄 환경요소
일상활동이론	- 동기가 부여된 가해자 - 보호능력의 부재 - 적절한 표적
범죄패턴이론	- 마디 - 경로 - 가장자리
의식이론	- 피해자 - 가해자 - 시공간 - 규제

표 2와 같이 데이터마이닝 기법을 적용하기 위해 활용되는 변수는 범죄자와 직, 간접적으로 연관된 환경정보, 범죄 행동 정보 등이 있으며, 가해자에 대한 특징 정보와 사건이 발생한 지역의 환경, 그리고 범행의 특징 및 발생한 범죄의 유형을 활용하여 데이터마이닝 기법을 적용할 수 있다.

표 2. 데이터마이닝에서 활용되는 변수

	데이터마이닝 활용 변수
Buczak, A.L. (2010), [12]	인종, 나이, 인구비율, 평균수입, 학력, 직업, 거주형태, 외국인비율, 인구밀도, 범죄유형 등
Nath, S.V. (2007), [4]	인종, 가해자인종, 가해자성별, 가해자나이, 피해자나이, 범행도구, 범죄유형 등
Dzemy diene, D. (2002), [13]	시간, 장소, 가해자나이, 가해자성별, 피해자나이, 피해자성별, 인구구조, 지역범죄율, 범죄유형 등

지금까지 살펴본 범죄학에서의 범죄 환경요소와 데이터마이닝에 활용되는 변수와의 관계는 표 3과 같이 유추할 수 있다. 인종, 성별 및 나이의 경우 가해자와 피해자와 직접 관련되어 있으며 객체의 특징을 대변하는 정보이며, 성별과 나이

표 3. 범죄 환경요소 및 데이터마이닝 활용변수 간 관계

		범죄 환경요소							
		보호능력의 부재	가해자	피해자	마디	경로	가장자리	시공간	규제
데이터마이닝 활용 변수	인종		○	○					
	성별	○	○	○					
	나이	○	○	○					
	인구비율				○			○	
	인구구조				○			○	
	평균수입				○			○	
	학력		○	○					
	직업		○	○	○	○	○		
	거주형태				○			○	
	지역범죄율				○				○
	시간							○	
	장소				○	○	○	○	
	범행도구		○						
	범죄유형								

는 피해자의 보호능력(Guardianship)과도 관련이 있다. 인구비율, 인구구조, 평균수입 및 거주형태는 지역(공간)의 인구학(Demography)적 특성을 보이므로 마디 및 시공간에 영향을 미친다 할 수 있다. 학력과 직업은 가해자 혹은 피해자와 관련된 특징 정보이며, 직업에 따라 주 활동 범위, 이동 경로 등이 달라지기 때문에 마디, 경로 및 가장자리와도 연관성이 있다. 지역범죄율은 주 거주 지역의 공간적인 환경과 개인 및 이웃의 생활패턴에 영향을 미치며 해당 지역의 범죄 예방에도 영향을 미치므로 규제의 역할을 한다고 볼 수 있다. 또한, 시간이나 장소는 마디, 경로 및 가장자리와 같은 시공간 요소와 관련이 있다. 범행도구의 경우, 가해자의 행동과 관련이 있다고 볼 수 있다. 이에 반해, 범죄예측 활용변수 중 범죄 유형은 다양한 범죄 환경요소와는 관련이 없으며 실제로 발생한 범죄에 관한 정보이다.

범죄 환경요소와 범죄예측 활용 변수 간의 관계에 비추어 볼 때 데이터마이닝은 범죄 환경구성요소와 범죄와의 관련성을 찾는 것이라 할 수 있으며, 새롭게 얻은 정보는 범죄 환경 요소에 대한 확실하지 않거나 기존에 알려지지 않은 새로운 정보라 할 수 있다. 데이터마이닝의 결과를 활용하여 상습범죄에 대한 여죄를 추적하거나 범죄자의 프로파일링에 적용하는 등 다양한 방면에서 실제 범죄수사에 활용될 수 있으며, 새로운 정보를 범죄 해결에 활용하는 것은 타당하다.

III. 데이터마이닝 기법

범죄의 패턴을 분석하고 특징을 추출하기 위해 활용되는 대표적인 데이터마이닝 기법으로 연관성분석(Association analysis), 군집분석(Cluster analysis), 분류분석(Classification analysis)이 있으며, 다양한 기법을 활용하여 범죄를 예측하고 범죄의 특징을 추출하는 연구가 진행되고 있다.

1. 연관성분석

연관성분석은 항목 간의 조건과 그에 따른 결과를 식으로 표현하는 연관성 규칙(패턴)을 발견하는 것을 의미하며, 범죄와 관련된 다양한 요인들 간 관계에 대한 정보를 얻기 위한 연구가 진행되고 있다[12][14]. 또한, 이와 같은 다양한 요 인간의 연관성을 활용하여 범행 유발 요인을 분석할 수 있다[15]. 데이터의 분석은 일반적으로 수학과 통계학의 확률과 기대치 개념을 기반으로 하나, 물리 및 수학에서 도출되는 규칙과 달리 100% 정확한 것이 아니다. 따라서 그 패턴이 얼마나 정확한지에 대한 성능척도가 중요하며, 패턴의 성능척도가 좋지 않더라도 범죄 해결을 위해 중요한 패턴일 수 있기 때문에 이를 쉽게 간과할 수 없다. 예를 들어, 살인범죄의 경우 발생 건수가 상대적으로 낮아 절도범죄 혹은 폭력범죄보다 패턴의 정확도가 낮을 수 있다. 그러나 살인범죄 패턴의 정확도가

낮다 하더라도 다른 범죄에 비해 위험성이 상대적으로 높기 때문에 쉽게 간과해서는 안 된다.

2. 군집분석

군집분석은 유사한 속성을 갖는 객체들을 묶어 전체의 객체들을 몇 개의 그룹 또는 군집(Cluster)으로 나누는 것을 의미한다. 군집분석은 연관성분석과 마찬가지로 범죄와 관련된 다양한 특징을 추출하는 데 활용되고 있으며[16][17], 군집의 특징정보를 활용하여 범죄예측에 활용할 수 있다[4][18]. 이 외에도 인구학적 및 사회학적 요소들을 고려하여 범죄의 패턴을 찾고자 하는 연구가 진행되고 있다[19]. 군집분석은 크게 계층적 방법(Hierarchical method)과 비계층적 방법(Non-hierarchical method)으로 구분된다. 계층적 방법은 사전에 군집 수 K 정하지 않고 단계적으로 군집분석을 진행하는 것이며, 비계층적 방법은 사전에 K를 정하고 각 객체는 K개의 군집 중 하나에 배정된다. 그러나 적절한 수의 K를 정하기는 쉽지 않기 때문에, 범죄의 특징을 잘 구분할 수 있는 K를 선정해야 한다. 또한, 군집분석을 위해서 유사성 척도를 사용하게 되는데, 데이터에 특성(연속형, 명목형)에 따라 적절한 유사성 척도를 선정해야 한다. 특히, 범죄 데이터의 경우 일반적으로 데이터는 연속형과 명목형 값이 혼합되어 있기 때문에 유사성 척도의 선정이 중요하다.

3. 분류분석

분류분석은 다수의 속성(Attribute) 또는 변수를 사전에 정의된 그룹(Class)으로 분류하는 것을 의미하며 범죄의 유형을 분류하거나[20], 분류된 결과를 활용하여 범죄를 예측하는 데 활용된다[21]. 또한 분류된 결과를 활용하여 재판과 관련된 의사결정을 하는데 활용하거나[22], 범죄자 프로파일링에 활용할 수 있다[23]. 분류분석은 객체를 몇 개의 그룹으로 구분한다는 점에서 분류분석은 군집분석과 유사하나, 분류분석은 지도학습(Supervised learning)이고 군집분석은 자율학습(Unsupervised learning)이라는 점에서 차이가 있다. 분류분석은 사용하고자 하는 목적 및 데이터의 특성에 따라 통계적 방법, 트리 기반 기법 혹은 비선형 최적화 기법을 적용할 수 있다.

IV. 사례연구

본 장에서는 실제로 우리나라에서 발생한 절도범죄 기록을 활용하여 앞 장에서 소개한 데이터마이닝 기법을 적용하였다.

연관성분석을 통해 범행 특징간의 관계를 분석할 수 있으며 분류분석을 통해 새로운 범죄의 유형을 분류 할 수 있다. 그러나 본 장에서는 군집분석을 통해 주위 환경 및 시공간적 요소와 범죄 특성이 어떤 관계가 있는지 알아보고, 이를 통해 데이터마이닝의 결과가 갖는 의미에 대해 살펴보고자 한다.

1. 활용 데이터

본 연구를 위해 국내에서 발생했던 절도범죄와 관련된 데이터를 수집하고 활용하였다. 데이터는 2010년부터 2013년까지 발생한 범죄를 대상으로 보호관찰소에 보관되고 있는 범죄 기록을 활용하였다. 범죄 기록은 평문으로 작성되어있었으며, 전문가 의견을 기반으로 정의한 변수를 활용하여 평문 기록을 데이터베이스화하였다. 이를 위해 일반특성, 범행예비, 범행방법, 피해자관계 등으로 구분하고 표 4에 정의된 변수를 포함한 총 65개의 변수를 활용하였으며, 발생한 범죄 건수가 가장 많은 절도 범죄 755건의 데이터를 수집하였다.

표 4. 데이터마이닝 입력 데이터

대분류	중분류	소분류
일반특성	성별	남자/여자
	가해자출생연도(나이)	()년도
	범행지역	()구/ ()동
	범행시간대	새벽/오전/ 오후/밤
	발생날짜(범행요일)	년/월/일

범행예비	사전대화유무	없음/있음
	가해자음주유무	없음/있음
	피해자음주유무	없음/있음
	접근미행유무	없음/있음

범행방법	손으로때림	없음/있음
	발로때림	없음/있음
	흉기로때림	없음/있음
	위협어부	없음/있음

사건정보	봉사시간	시간
	수감시간	시간
	보호관찰기간	개월
	전과횟수	횟수
...

일반특성으로 분류된 변수는 범행과 관련된 일반적인 정보

와 시공간 요소를 포함한 주위 환경정보를 의미한다. 범행에 비로 분류된 변수는 범행 전 가해자와 피해자와 관련된 사건 행동정보이며, 범행방법은 가해자와 관련된 행동정보를 의미한다. 사건정보는 가해자의 범행에 따른 처벌과 관련된 정보를 이며, 범죄학에서의 범죄 환경요소와는 관련이 없다.

각각의 범죄 사건에 관해서 모든 변수의 정보를 포함하고 있지 않았기 때문에 데이터의 손실이 발생하였으나 그 수가 많지 않아 손실이 있는 사건은 본 연구에서 제외하였다. 또한, “가해자출생연도”는 가해자의 “나이”로 변환하였으며, “발생날짜”의 경우 140302와 같이 년/월/일로 입력된 데이터를 요일별 특성을 도출하기 위해 “범행요일”로 변환하여 사용하였다.

2. 실험 및 결과

본 연구에서는 데이터마이닝 기법 중 군집분석을 적용하였다. 각 변수의 값이 이분형 혹은 명목형임을 고려하여 유사성 척도를 선정하였다. 또한, 사건정보는 군집분석에 따른 군집 간 비교분석을 위해 활용하였으며, 유사성 척도 계산에서는 제외하였다. 군집분석에 따른 군집 수와 군집의 기본 통계 정보는 표 5와 같다.

표 5. 군집분석 기본 통계 분석

군집 ID	군집 사건수	평균 봉사 시간 (시간)	평균 수강 시간 (시간)	평균 보호관찰 기간 (개월)	평균 전과 횟수 (횟수)
군집 1	101	45.15	5.35	16.34	4.22
군집 2	556	28.25	6.41	15.97	3.61
군집 3	11	18.18	7.27	9.55	3.09
군집 4	87	24.37	4.83	14.73	3.36
합계	755	-	-	-	-

군집 1의 경우, 평균 전과횟수가 가장 많았으며 군집 3의 경우가 평균 전과횟수가 가장 적었다. 평균 전과횟수가 많을 수록 봉사시간과 보호관찰기간이 긴 것을 알 수 있으며, 수강 시간은 전과횟수와 관련성이 없는 것으로 보인다. 또한, 각각의 군집의 속성을 비교하였으며 그 결과는 그림 3과 같다.

군집 1과 2의 경우, 사건은 주로 새벽에 남성을 대상으로 발생하였으며 군집 1의 사건들은 유흥업 및 길거리에서 주로 발생했지만 군집 2는 공공장소와 길거리에서 발생한 사건의 비율이 높음을 알 수 있다. 군집 3과 4의 경우, 밤에 발생한 사건의 비율이 군집 1과 2에 비해 높으며 특히 군집 3의 경우, 여성을 대상으로 길거리에서 주로 사건이 발생하였다. 군집 3과 4의 범행예비에 따른 사건 비율은 군집 1과 2와 비교

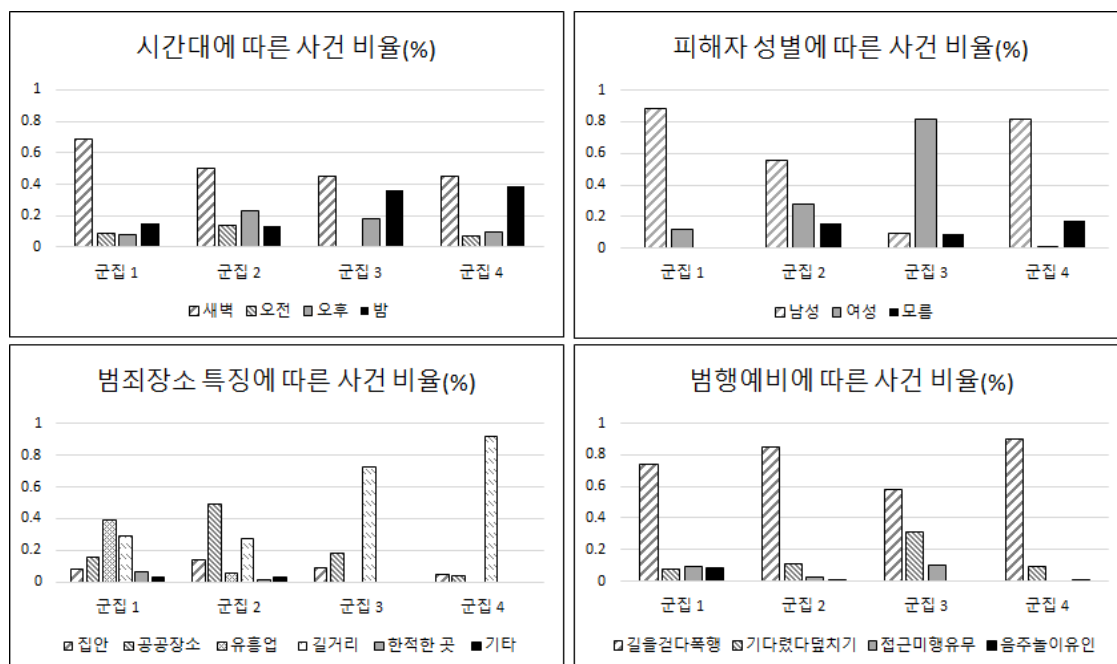


그림 3. 군집 속성에 따른 비교 예시

했을 때, 큰 차이가 없으며 주로 가해자는 주로 길을 걷다 피해자를 폭행하였음을 알 수 있다. 또한, 주로 여성을 대상으로 한 군집 3의 경우, 다른 군집에 비해 가해자가 피해자를 접근 혹은 미행한 사건의 비율이 높음을 알 수 있다.

이처럼 군집분석을 통해 얻은 군집의 특징은 범죄학이론의 범죄 환경요소와 범죄와의 관련성을 보여준다. 각 군집별로 범죄의 특성이 다르며 각 군집의 특성은 범죄와 관련된 패턴이라 할 수 있다. 특히, 각 패턴은 군집분석에 활용된 변수에 따라 결정되며 이때 활용된 변수는 범죄학에서의 범죄 환경요소이다. 예를 들어, 군집분석에서 활용된 변수인 시간대와 범죄장소는 시공간 정보이며, 피해자 성별 및 범행예비행동은 피해자와 가해자의 특징이다. 따라서, 군집분석을 통해 얻은 정보는 범죄학에서의 발생한 범죄와 범죄 환경과의 관계라 할 수 있다. 즉, 군집분석과 같이 데이터마이닝을 통해 얻을 수 있는 정보는 범죄 환경요소와 범죄와의 관계를 보이는 것이며, 위 결과를 활용하여 범죄학관점에서 범죄 예측 및 예방하는 것은 유의미하다.

V. 결론

데이터마이닝은 데이터로부터 유용한 정보를 추출하는 과정을 의미한다. 추출한 정보가 실제로 의미가 있고 실제로 범죄 사건 해결을 위한 의사결정에 도움을 주기 위해서는 정보의 타당성에 대한 논의가 필요하며 어떤 의미가 있는지에 대한 분석이 필요하다.

데이터마이닝이 범죄 해결에 실제로 도움을 줄 수 있는지 알아보기 위해 먼저, 범죄학의 범죄 발생 환경에 대해 알아보고 범죄 환경요소와 데이터마이닝에 활용되는 변수가 실제로 유사함을 보였다. 또한, 국내에서 실제로 발생한 절도 범죄 데이터를 활용하여 주위 환경정보와 범죄 특성 간의 관계를 보임으로, 주위 환경과 범죄의 특징이 실제로 관계가 있음을 보였다.

군집분석을 통해 추출된 주위 환경과 범죄 특성 간의 관계는 범죄학의 주위 환경과 가해자의 관계와 비교해 볼 수 있다. 다양한 범죄학 이론들은 가해자의 행동이 주위 환경과 같은 시공간 요소들에 영향을 받는다고 주장하며 요소에 따라 가해자의 특성이 변할 수 있음을 염두에 두고 있다. 실제로 군집분석을 통해서 시간대 및 장소의 특징에 따라 피해자의 성별과 가해자의 범행예비 특성에 차이가 있음을 보였으며, 시간대 및 장소에 따라 다른 특징이 있는 범죄가 발생한다는 정보를 얻을 수 있었다. 즉, 군집분석을 통해 얻은 정보는 가해자와 주위 환경과의 관계에 관한 정보이며 데이터마이닝을 범죄

예측에 활용한다는 것은 범죄를 구성하고 있는 다양한 요소들을 활용하여 새로운 정보를 추출하는 것이다. 따라서 기존에 알지 못했던 새로운 정보를 얻는다는 것은 범죄를 구성하는 요소들 간의 새로운 관계를 밝히는 것이므로 유의미하며 범죄학적 특징을 잘 반영한다 할 수 있다.

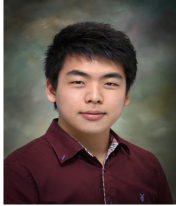
본 연구는 범죄예측에서 데이터마이닝을 활용한다는 것이 어떤 의미인지 알아보았다는 점에서 의의를 가진다. 그러나 데이터마이닝에서 활용되는 변수와 범죄 환경요소 간의 관계를 체계적인 방법론을 활용하여 보이지 못한 점에 한계가 있다. 또한, 충분한 수의 데이터를 수집하지 못했기 때문에 데이터마이닝을 통해 추출된 정보가 구체적으로 범죄 환경요소와 어떤 관계가 있는지 보이지 못하였다. 앞으로 충분한 수의 데이터를 수집하고 절도 범죄가 아닌 5대 범죄 중 범죄의 특성이 뚜렷한 강도 및 살인을 대상으로 사례연구를 진행한다면, 가해자와 주위 환경요소들 간의 관계를 더 명확히 보일 수 있을 것으로 예상된다.

참고문헌

- [1] C. Shu, A. Hampapur, M. Lu, L. Brown, J. Connell, A. Senior, and Y. Tian, "IBM Smart Surveillance System(S3): A Open and Extensible Framework for Event based Surveillance," IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 318-323, Como, Italy, September 2005.
- [2] S. Chainey, L. Thompson, and S. Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime," Security Journal, Vol. 21, pp. 4-28, February 2008.
- [3] D. Brown, and R. Oxford, "Data Mining Time Series with Applications to Crime Analysis," IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 1453-1459, Tucson, USA, October 2001.
- [4] S. Nath, "Advances and Innovation in Systems, Computing Science and Software Engineering," Springer, pp. 405-409, 2007.
- [5] H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime Data Mining: A General Framework and Some Examples," Computer, Vol. 37, No. 4, pp. 50-56, April 2004.

- [6] V. Grover, R. Adderley, and M. Bramer, "Applications and Innovations in Intelligent Systems XIV," Springer London, pp. 233-237, 2007.
- [7] M. Keyvanpour, M. Javideh, and M. Ebrahimi, "Detecting and Investigating Crime by Means of Data Mining: A General Crime Matching Framework," *Procedia Computer Science*, Vol. 3, pp. 872-880, February 2011.
- [8] P. Thongatae, and S. Srisuk, "An Analysis of Data Mining Application in Crime Domain," *IEEE International Conference on Computer and Information Technology Workshops*, pp. 122-126, Sydney, Australia, July 2008.
- [9] L. Cohen, and M. Felson, "Social Change and Crime Rate Trends: A Route Activity Approach," *American Sociological Review*, Vol. 44, No. 4, pp. 588-608, August 1979.
- [10] M. Felson, and R. Clarke, "Opportunity Makes the Thief: Practical Theory for Crime Prevention," *Police Research Series Paper 98*, Home Office, pp. 4-8, 1998.
- [11] P. Brantingham, and P. Brantingham, "Environmental Criminology," *Waveland Press Inc*, pp. 27-54, 1991.
- [12] A. Buczak, and C. Gifford, "Fuzzy Association Rule Mining for Community Crime Pattern Discovery," *ACM SIGKDD Workshops on Intelligence and Security Informatics Article*, No. 2, New York, USA, July 2010.
- [13] D. Dzemydiene, and V. Rudzkiene, "Multiple Regression Analysis in Crime Pattern Warehouse for Decision Support," *Proceedings of the 13th International Conference on Database and Expert Systems Applications*, pp. 249-258, Aix-en-Provence, France, September 2002.
- [14] V. Ng, S. Chan, D. Lau, and C. Ying, "Incremental Mining for Temporal Association Rules for Crime Pattern Discoveries," *Proceedings of the Eighteenth Conference on Australasian database*, Vol. 63, pp. 123-132, Darlinghurst, Australia, March 2007.
- [15] B. R. Mednick, R. L. Baker, and L. E. Carothers, "Patterns of Family Instability and Crime: The Association of Timing of the Family's Disruption with Subsequent Adolescent and Young Adult Criminality", *Journal of Youth and Adolescence*, Vol. 19, No. 3, June 1990.
- [16] H. Liu and D. E. Brown, "Criminal Incident Prediction Using a Point-pattern-based Density Model", *International Journal of Forecasting*, Vol. 19, No. 4, pp. 603-622, December 2003.
- [17] T. Nakaya, and K. Yano, "Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics," *Transaction in GIS*, Vol. 14, No. 3, pp. 223-239, June 2010.
- [18] B. Chandra, M. Gupta, and M. Gupta, "A Multivariate Time Series Clustering Approach for Crime Trends Prediction," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 892-896, Singapore, October 2008.
- [19] M. Carglia, R. Haining, and P. Wiles, "A Comparative Evaluation of Approaches to Urban Crime Pattern Analysis", *Urban Studies*, Vol. 37, No. 4, pp. 711-729, April 2000.
- [20] R. Jamieson, L. Land, D. Winchester, G. Stephens, A. Steel, A. Maurushat, and R. Sarre, "Addressing Identity Crime in Crime Management Information Systems: Definitions, classification, and empirics," *Computer Law & Security Review*, Vol. 28, No. 4, pp. 381-395, August 2012.
- [21] A. Nasridinov, S. Ihm, and Y. Park, "Information Technology Convergence," *Springer Netherlands*, Vol. 253, pp. 531-538, 2013.
- [22] D. M. Gottfredson, "Prediction and Classification in Criminal Justice Decision Making", *Crime and Justice*, Vol. 9, pp. 1-20, 1987.
- [23] J. E. Douglas, R. K. Ressler, and C. R. Hartman, "Criminal Profiling from Crime Scene Analysis", *Behavioral Science and the Law*, Vol. 4, No. 4, pp. 401-421, Autumn 1986.

저 자 소 개



방 승 환

2013: 경희대학교

산업경영공과 공학사.

현 재: 포항공과대학교

산업경영공학과 통합과정.

관심분야: 산업경영공학

Email : seunghwanb@postech.ac.kr



김 태 훈

2012: 포항공과대학교

산업경영공과 공학사.

현 재: 포항공과대학교

산업경영공학과 통합과정.

관심분야: 산업경영공학

Email : taehun_kim@postech.ac.kr



조 현 보

1986: 서울대학교

산업공학 공학사.

1988: 서울대학교

산업공학 공학석사.

1993: Texas A&M University

산업공학 공학박사

현 재: 포항공과대학교

산업경영공학과 교수

관심분야: 산업경영공학

Email : hcho@postech.ac.kr