

# Industrial Safety Risk Analysis Using Spatial Analytics and Data Mining

Kyeongseok Ko · Jaekyung Yang<sup>†</sup>

Dept. of Industrial and Information Systems Engineering, Chonbuk National University

## 공간분석 · 데이터마이닝 융합방법론을 통한 산업안전 취약지 등급화 방안

고경석 · 양재경<sup>†</sup>

전북대학교 산업정보시스템공학과

The mortality rate in industrial accidents in South Korea was 11 per 100,000 workers in 2015. It's five times higher than the OECD average. Economic losses due to industrial accidents continue to grow, reaching 19 trillion won much more than natural disaster losses equivalent to 1.1 trillion won. It requires fundamental changes according to industrial safety management. In this study, We classified the risk of accidents in industrial complex of Ulju-gun using spatial analytics and data mining. We collected 119 data on accident data, factory characteristics data, company information such as sales amount, capital stock, building information, weather information, official land price, etc. Through the pre-processing and data convergence process, the analysis dataset was constructed. Then we conducted geographically weighted regression with spatial factors affecting fire incidents and calculated the risk of fire accidents with analytical model for combining Boosting and CART (Classification and Regression Tree). We drew the main factors that affect the fire accident. The drawn main factors are deterioration of buildings, capital stock, employee number, officially assessed land price and height of building. Finally the predicted accident rates were divided into four class (risk category-alert, hazard, caution, and attention) with Jenks Natural Breaks Classification. It is divided by seeking to minimize each class's average deviation from the class mean, while maximizing each class's deviation from the means of the other groups. As the analysis results were also visualized on maps, the danger zone can be intuitively checked. It is judged to be available in different policy decisions for different types, such as those used by different types of risk ratings.

**Keywords** : Industrial Safety, Spatial Analytics, Geographically Weighted Regression, Data Mining

### 1. 서론

각종 안전보건 대책(산재 예방 5개년 계획, 중대 재해 예방대책 등)의 마련·시행으로 재해율 지표는 꾸준히 개선되고 있지만, 사망사고만인 비율은 여전히 선진국보다

2~4배 가량 높다. 그리고 산업재해로 인한 경제적 손실액도 지속적으로 증가하여 자연재해 손실액의 110배에 해당하는 19조 원에 달하고 있다. 단기적 성과중심의 경영 확산으로 사업장의 안전보건체제가 안착되지 못하고, 사업주와 근로자의 안전의식이 여전히 낮은 상황이다. 현장과 괴리된 법제, 공급자 중심의 지원 사업 추진 등으로 안전보건 환경 변화에 주도적 대응이 곤란한 실정이다[16].

우리나라는 1970년대 이후 성장 위주의 경제성장 전

Received 4 September 2017; Finally Revised 9 November 2017;  
Accepted 10 November 2017

<sup>†</sup> Corresponding Author : jkyang@jbnu.ac.kr

락을 선택하면서 근로현장에서 발생하는 산업재해를 산업발전에 수반하여 어쩔 수 없이 발생하는 것으로 인식해왔다. 더욱이 사회와 기업이 산업재해가 단순히 근로자 개인의 부주의에 기인한다는 점만을 강조함에 따라 산업재해 발생의 본질을 왜곡하여 산업재해의 문제해결을 더욱 어렵게 만들어왔다. 산업구조 생산방식의 변화, 안전보건 취약계층의 증가, 재해요인의 다양화는 안전정책의 초점과 접근방식의 변화를 요구하고 있다[14].

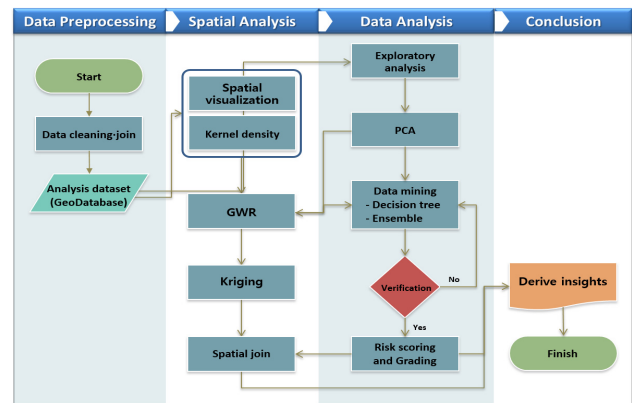
'90년대부터 서비스업의 비중이 지속적으로 증가하고 제조업의 비중은 계속해서 감소하는 산업구조의 변화가 진행 중이다. 또한, 유해·위험 작업을 직접 수행하지 않고 도급을 하는 아웃소싱의 일반화, 저출산·고령화와 외국인 근로자의 증가, 50인 미만 소규모 사업장이 지속적으로 증가 추세이다. 그러므로 노동력의 양적·질적인 변화에 맞추어 안전점검 및 교육정책을 재구성할 필요가 있다. 다양한 환경, 위험요인에 맞는 정책을 위해서는 산업의 특성에 따른 선제적이고 효과적인 산업 안전정책 대응체계의 구축이 필요하다[16].

본 연구의 목적은 산업단지에서 일어나는 사고를 중심으로 사고에 영향을 끼치는 데이터를 수집하여 공간분석 및 데이터마이닝을 수행하여 산업단지의 취약지를 도출하는데 있다. 이 분석결과를 위험등급별, 지역별 맞춤형 산업안전사고 예방 및 감소를 위한 산업재해 예방정책 수립에 활용할 수 있다. 본 연구를 위해 대규모 화학처리 시설과 노후화된 산업단지가 밀집된 울주군을 대상으로 선정하여 사고의 발생 위치, 사고유형, 사고 발생시간 및 사고 발생업체의 업종·노후연수·근로자 수·외국인노동자 수, 건축물 정보 등의 데이터를 활용하여 산업단지 내에 업종별, 사고유형별 사고특징을 탐색하고, 사고 요인을 도출한다. 사고 발생예측 분석을 통해 위험지수를 개발하여 위험도에 따른 “위험, 경계, 주의, 관심” 4가지로 등급화 하였다. 또한, 분석결과를 지도로 시각화하였다[13]. 기존에 여러 공간분석을 이용한 연구가 있었는데 공간정보기반의 화산재해대응시스템 개발 및 활용[11], 상황인지모델을 이용한 GIS 기반의 대기오염 정보 시스템 연구[12], 침수지도 관리를 위한 벡터형 공간정보 표준 기반의 메타데이터 설계연구[17] 등이 있었으나 이들 연구는 공간분석 방법론만을 이용한 연구라는 점에서 차이가 있다. 또한 본 연구에서 공간분석과 융합되는 데이터마이닝은 많은 데이터에서 숨겨져 있던 새롭고 의미 있는 정보를 찾는 데 사용되는 기술로써 많은 연구가 있었다. 예를 들어 기계적 모터 고장진단을 위한 머신러닝 기법 연구[7], SVM과 인공신경망을 이용한 침입탐지 연구[6], 데이터 유출 참지를 위한 이상 행위 탐지 방법 연구[15], 효과적인 에너지 관리를 위한 데이터마이닝 모델 연구[8] 등 이 있다. 이러한 공간분석과 데이터마이닝 융합방법론을 적용한 예측 모델을 활용해서 앞으로 발생할

산업단지 내 사고에 선제적인 대응을 하는데 목적이 있다.

## 2. 연구방법

본 연구의 과정은 <Figure 1>에 도식화 한 바와 같이 크게 데이터 전처리, 공간분석, 데이터 분석, 결론 도출 순서로 진행하였다. 데이터 전처리에서는 데이터의 정제, 융합하여 분석 데이터셋을 구축하였다. 공간분석에서는 산업안전사고 유형별 분포, 공장 특징별 분포를 시각화하고, 커널밀도 분석을 수행했다. 데이터 분석에서는 수많은 변수 중에서 종속변수에 영향을 끼치는 주요 요인을 추출하는 주성분 분석을 통해 사고 주요 요인을 도출하였고, 도출된 요인을 활용하여 공간적 이질성, 공간적 영향력을 고려한 지리적 가중회귀분석(GWR : Graphically Weighted Regression)을 통해 공간 단위별 예측값을 도출하였다. 그리고 GWR의 예측값과 관심 있는 지점에서의 특성치를 알기 위해 이미 그 값을 알고 있는 주위의 값들의 선형 조합으로 그 값을 예측하는 기법인 크리깅을 통한 확률추정값을 비교하여 검증하였다. GWR의 분석결과인 예측값을 데이터마이닝의 분석 가중치로 추가하여 CART 분석에 부스팅기법이 결합된 분석모델을 통해 지역별 위험도를 점수화하고 등급화하였다. 등급화한 결과를 공간적으로 시각화하였고, 분석결과를 통해 울주군 산업단지 내 화재사고 취약지를 도출하였다.



<Figure 1> Research Process

## 3. 공간분석 및 데이터마이닝

### 3.1 데이터목록

원시데이터는 다음 <Table 1>과 같이 사고데이터인 119 구조출동 정보, 공장특성 데이터, 매출액, 자본금 등의 기업 정보, 건축물 정보, 기상정보, 공시지가 등으로 구성하였다.

〈Table 1〉 Data List

Data	Source	Main attributes	Spatial unit
119 rescue operations	Ulsan fire department	Address, date, time, kind, contents, etc.	Parcel
Factory Info.	FactoryOn	category of business, Scale, worker, etc.	Parcel
Building ledger	EAIS	Usage, main structure, start date, etc.	Building
Weather Info.	Meteorological office	Temperatures, Precipitation, etc.	Observatory
Corporate info.	worknet	Sales, Capital, etc.	Parcel
land price	Ulju-gun	Official land price	Parcel
Building info.	NSIC	[Spatial data]Address, Area, Building polygon	Building
Continuous cadastral map	Ulju-gun	[Spatial data]	-
Administrative district	Ulju-gun	[Spatial data]	-
Aerial imaging	Ulju-gun	[Spatial data]	-
Industrial complex	Ulju-gun	[Spatial data]	-

공간정보는 “지리적인 위치 및 위상 관계를 나타내는 도형정보와 공간상의 사물에 대한 특징을 나타내는 속성정보의 집합체”로 정의할 수 있다. 도형정보는 지도확상의 좌표체계에 따라 좌표값을 기준으로 점, 선, 면 등의 형태로 표현되고, 속성정보는 아라비아 숫자나 텍스트 기호 등으로 표현된다. 특히, 공간정보를 입력할 때는 정보의 형상을 입력해야 하며 그 표현방법으로 벡터 데이터와 래스터 데이터의 형태가 있다. 공간정보의 가장 큰 특징은 도형정보와 속성정보가 1:1로 연결되어 관리된다는 것이다.

### 3.2 데이터 전처리 및 융합

수집된 일반 속성정보 5종, 공간정보 5종의 데이터를 융합하기 위해 일반 속성정보를 공간화, 즉 지오코딩 과정을 거쳐서 모두 공간정보로 변환하고, 공간조인과정을 거쳐서 모든 수집된 데이터를 격자 단위로 융합하였다.

2011년 1월~2015년 9월까지 울주군의 119 출동데이터를 활용하여 결측치 제거 및 정제 등 전처리 후 411건 사고데이터를 만들고, 위치와 사고일자를 기준으로 기상데이터와 조인(Join)하였다. 울주군 내 운영 중인 1,353건의 공장 속성데이터를 지오코딩 과정을 거쳐 공간화 하였다. 공장 속성 중 주소정보를 공간상의 x, y좌표로 변환해주는 지오코딩 툴을 활용하여 공장의 좌표 정보를 획득하고 공간화 및 시각화는 ArcGIS라는 공간정보를 편집, 가공, 분석할 수 있는 툴을 활용하였다.

울주군에 건축물 속성데이터(159,425건)를 조인(Join)을 통해서 공장 속성데이터 및 사고 데이터와 융합하였다.

울주군 전역을 대상으로 500m×500m 격자를 생성하고 사고 데이터와 기상 데이터의 융합 데이터를 격자를 기준

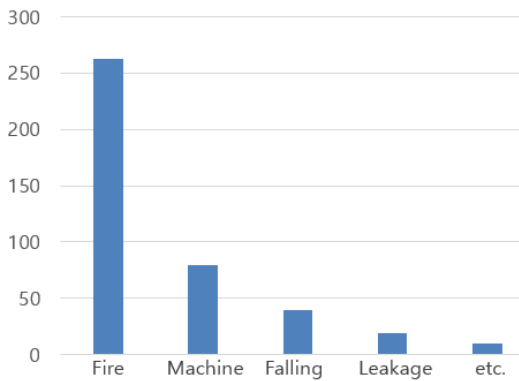
으로 공간조인(Spatial Join)하여 사고 융합 데이터셋을 만들었다. 건축물대장, 공장특성 데이터를 PNU(Parcel Number Unit)로 테이블 조인(Join) 한 후 사고 융합데이터셋(격자)에 건축물대장과 공장특성데이터를 융합하여 통합 데이터셋을 구성하였다. 최종 분석 데이터셋은 울주군이라는 공간적 범위를 격자 단위로 나눠 산지, 주거지역을 제외한 공장이 존재하는 지역의 119개의 격자(records)로 만들었다. 속성정보는 <Table 2>와 같이 사고건수의 합, 건축노후연수의 합, 공장면적의 합, 종사자 수의 합, 건물 높이의 합, 건물 면적의 합, 공시지가의 합, 자본금의 합, 매출액의 합 등의 45개 속성으로 구성했다[18].

〈Table 2〉 Final Analysis Data Specification

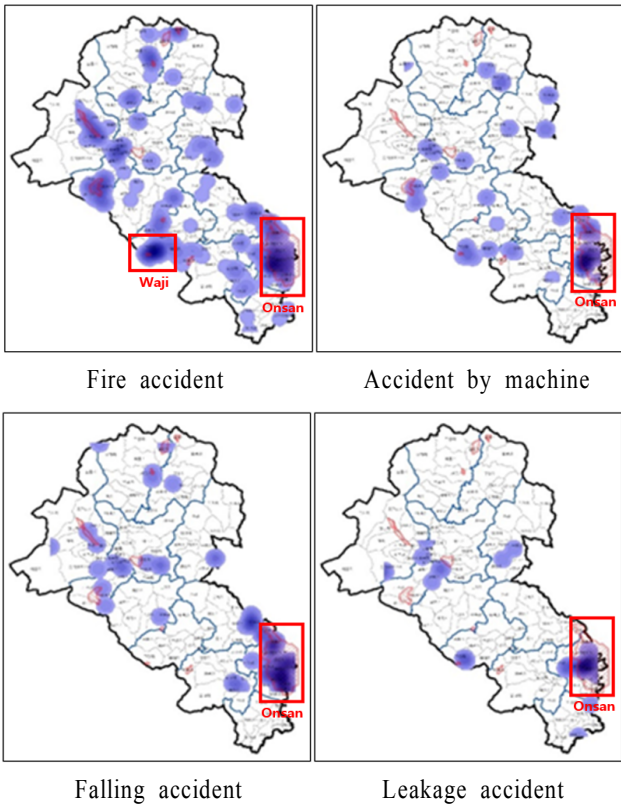
Field name	Explanation	Data type
OBJECTID	ID	Object ID
Shape	Geometry	Geometry
AvgAvgTmp	Average temperature	Numeric
SumDayRf	Daily precipitation	Numeric
AvgMxWs	Maximum wind speed	Numeric
SumAvgWs	Average wind speed	Numeric
SumAcc	Sum of fire accidents	Numeric
SumAccLe	Sum of undercarriage accidents	Numeric
SumAccMch	Sum of machine accidents	Numeric
SumAccFa	Sum of crashes	Numeric
SumAccEtc	Sum of other accidents	Numeric
AvgFacCnt	Number of Factory	Numeric
SumDngAr	Sum of Dong area	Numeric
SumHeight	Sum of building height	Numeric
SumFloors	Sum of floors	Numeric
SumPltAr	Sum of land area	Numeric
SumArcAr	Sum of building area	Numeric
SumTotAr	Sum of floor area	Numeric
SumEmpMA	Sum of male workers	Numeric
SumEmpWA	Sum of female workers	Numeric
SumEmpA	Sum of workers	Numeric
SumCapi	Sum of capital	Numeric
SumSale	Sum of sales	Numeric
SumJiga	Sum of land price	Numeric
SumOldYr	Sum of building old years	Numeric
SumBdAr	Sum of building area(FactoryOn)	Numeric
SumFArea	Sum of factory area	Numeric
SumTypAll	Sum of all factories	Numeric
SumTypFo	Sum of food companies	Numeric
SumTypCth	Sum of textile apparel companies	Numeric
SumTypWo	Sum of wood paper companies	Numeric
SumTypPtr	Number of petrochemical companies	Numeric
SumTypNmt	Number of non-metallic companies	Numeric
SumTypItr	Number of Steel Companies	Numeric
SumTypMch	Number of machinery manufacturers	Numeric
SumTypEl	Number of electronic companies	Numeric
SumTypTr	Number of transportation companies	Numeric
AvgTypEtc	Average number of other companies	Numeric
Shape Le	Total length of building shape	Numeric
SBC RAT	Sum of Coverage ratio	Numeric
SVL RAT	Sum of Floor area ratio	Numeric
SumFMEmp	Foreign male worker sum	Numeric
SumFWEmp	Foreign female worker sum	Numeric
SumFEmp	Foreign worker sum	Numeric
Shape Area	Total area of building shape	Numeric

### 3.3 탐색적 분석 및 공간 시각화

산업안전사고 분포를 공간시각화 해보면, 전체사고와 화재사고 모두 온산 국가산업단지, 와지 일반산업단지에 집중적으로 발생하고 있는 것으로 나타났다. 사고유형별 발생빈도를 분석해보면, <Figure 2>와 같이 화재사고, 기계(칼림끼임)사고, 추락사고, 누수누출사고, 기타사고의 5가지 유형으로 사고 분류할 수 있고, 화재사고가 전체사고 중 65%를 차지하여 상대적으로 가장 많이 발생하는 것을 볼 수 있다.



<Figure 2> Accident Type



<Figure 3> Kernel Density by Accident Type

사고 유형별 각기 다른 요인들이 사고에 영향을 끼치기 때문에 전체사고 중 가장 많은 비중을 차지하는 화재사고에 대하여 분석을 하였다. 사고유형별 분포를 <Figure 3>과 같이 살펴보면 화재사고는 울주군 전체에 넓게 분포하고, 특징적으로 와지 산업단지와 온산 국가산업단지의 오른쪽에 집중 분포하고 있고, 추락 및 누수·누출사고는 온산 국가산업단지에 집중적으로 분포하고 있는 것을 알 수 있다.

공장을 업종별로 나누어 보면, 기계, 석유화학, 운송장비, 전기·전자, 목재 종이출판, 음식료, 비금속 소재, 철강, 섬유·의복, 기타의 10가지 업종으로 분류할 수 있고, 이 중 전체공장의 30%를 차지하는 기계업종이 가장 큰 비중을 차지하고 있다. 전체 공장 중 가장 많은 비중을 차지하는 기계업종에 대하여 분석 수행하였다.

### 3.4 주성분 분석

사고 주요인자로 <Table 3>과 같이 자본금, 매출, 종사자 수, 건물 높이, 노후연수, 지가의 요인이 도출되었다.

<Table 3> PCA Result

	Components					
	1	2	3	4	5	6
AVG_TEMP	.037	.105	.001	-.074	.821	-.200
DAY_RF	-.117	.204	-.109	-.002	.499	.533
AVG_WS	.212	-.097	.038	-.017	-.117	.809
JIGA	.028	-.303	.062	.176	.508	.158
ZForn_sum	-.009	.155	-.245	.694	.092	-.142
ZCAPI	.760	-.183	.282	.070	.047	-.032
ZSALE	.912	-.081	-.086	-.022	.026	.079
ZEMP_TOT	.909	.024	-.026	-.010	.003	.082
ZF_AREA_SU	.815	-.142	.108	-.069	.010	.062
GRND_FLR	.083	.188	.794	.178	.019	.117
UGRND_FLR	.127	.103	.710	-.195	-.004	-.115
BC_RAT	-.097	.941	.121	-.008	.001	-.011
VL_RAT	-.098	.937	.187	-.026	.010	-.002
VL_RAT_TOT	.575	.090	.131	.293	-.078	.025
OLD_YY	-.106	.244	-.197	-.728	.038	-.125

- Extraction method : PCA  
 - Rotation : Kaiser Berimax with normalization

### 3.5 지리적 가중회귀분석

Moran's I 공간 자기 상관성 계수는 공간 자기 상관을 파악하기 위한 유용한 측정 척도로, 인접해있는 공간 단위들이 갖는 값을 비교하여 계수를 산출하게 된다. 인접한

공간 단위들이 전체 연구 지역에 걸쳐 유사한 값을 갖는 경우 Moran's I 계수는 높은 '정적(+)' 공간상관을 갖게 되고 그 반대의 경우 '부적(-)' 공간상관을 갖게 된다. Moran's I 계수는 다음의 식 (1)으로 표현될 수 있다[1].

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2} \quad (1)$$

위 식에서 공간가중치 매트릭스(spatial weights matrix) W의 구성인자인  $w_{ij}$ 는 공간 단위 i와 j가 서로 이웃하고 있으면 1, 그렇지 않으면 0의 값을 갖는다[2]. 공간 단위 i의 값은  $x_i$ 이며,  $\bar{x}$ 는 변수 x의 평균값이고, n은 관측개체의 수이다. Moran's I 계수는 -1에서부터 +1까지의 값을 갖는다.

사고(종속변수)에 영향을 미치는 요인별(독립변수)로 지리적 가중 회귀분석을 수행 위한 논리적 당위성을 확보하였다.

<Table 4> Moran's I Result

Independent variable	Moran's I	P value
Land price	0.2532	0.0001
Number of worker	0.0382	0.0034
Height	0.3961	0.0001
Capital	0.1622	0.0200
building old years	0.1293	0.0001

<Table 4>에서 Moran's I 계수로 주요 요인이 공간적 자기 상관이 있음을 확인하고 지리적 가중회귀분석(GWR)을 수행하였다. GWR의 경우 회귀모델을 사용한다는 점에서 기존 모형들의 장점이 적용됨과 동시에 지역별로 각각 다른 계수의 추정이 가능하여 기존 모형들로는 파악하기 어려운 공간적 이질성을 확인할 수 있다[5, 9]. GWR 결과를 GIS맵을 통해 제시함으로써 보다 시각적으로 데이터들의 공간적 상호작용을 파악할 수 있다[3].

전체 업종과 전체 사고로 분석을 해 본 결과보다 업종별, 사고유형별로 분류하여 분석하는 것이 정확도가 높게 나타나 가장 많이 분포하고 있는 업종인 기계업종과 화재 사고에 대한 분석 범위를 결정하였다. 다시 말해, 기계업종에 대한 화재사고데이터만을 학습데이터로 사용하였다.

자본금, 종사자 수, 노후연수, 지가, 건물 높이를 독립변수로 설정하고 종속변수는 화재사고 건수로 설정한 후 지리적 가중 회귀분석을 수행하였다. <Table 5>는 그 결과이다.

GWR은 지리적 이질성을 고려한 회귀식으로 지역마다 회귀식이 도출되었고, 대표적인 회귀식을 살펴보면, "화재 사고건수(y) = 건물높이×0.047825+자본금×2.913595+종사자

<Table 5> GWR Result

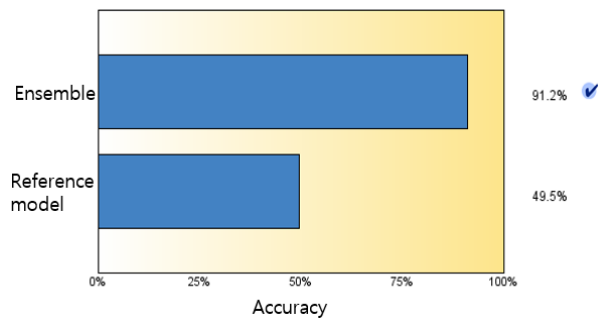
Name	Variable	Definition
Bandwidth	14188.506618	
ResidualSquares	233.363132	
EffectiveNumber	14.001621	
Sigma	1.490819	
AICc	455.80409	
R2	0.682307	
R2 Adjusted	0.642968	
Dependent Field	0	SumAccAll
Explanatory Field	1	ZSumJiga
Explanatory Field	2	ZSumEmpT
Explanatory Field	3	ZSumSale
Explanatory Field	4	ZSumOldYr
Explanatory Field	5	SumHeight

수×0.350813+지가×0.070182+노후연수×0.120898+0.470544"로 나타났다. GWR 분석결과와 RMSE(Root Mean Square Error)는 1.10(건)으로 나타났다.

모델의 설명력을 나타내는 Adjusted R<sup>2</sup>가 0.64를 나타내고 공간적으로 표현하면 0.52~0.68 사이에 분포함을 알 수 있다.

### 3.6 CART-부스팅 분석

GWR의 예측값을 분석가중치로 활용하여 의사결정트리 기법 중 하나인 CART(Classification and Regression Trees) 분석에 예측력을 개선하기 위한 앙상블(부스팅) 학습이 병합된 분석을 수행했다[10]. 종속변수가 화재사고 건수의 합으로 연속형 변수이고, 독립변수는 매출액, 용적률, 면적, 평균기온, 강수량으로 설정하고 부스팅과 CART 분석이 결합한 분석 모델을 통해 부모 노드로부터 자식 노드를 2개만 형성하여 분석해 나가며 화재사고 위험도를 스코어링했다. 데이터를 훈련데이터, 테스트데이터의 비율을 80:20으로 나눠 훈련데이터로 모델을 만들고 테스트 데이터를 통해서 검증을 수행했다.



<Figure 4> Ensemble Model Accuracy Comparison



CART 분석결과 <Figure 4>와 같이 Tree의 깊이도 알고 정확도가 49%로 낮아 분석결과로 채택할 수 없어서 가장 보편적인 부스팅 알고리즘인 Adaboost 기법 활용하여 91.2% 정확도의 모델을 개발하였다.

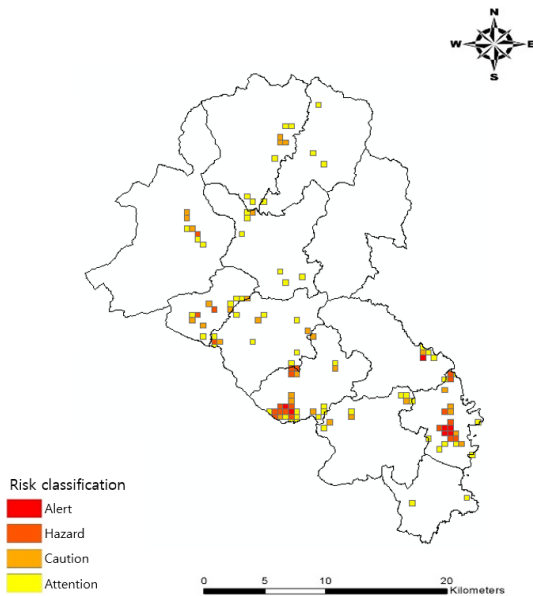
테스트 결과 평균절대오차 0.92(건), 예측값과 실제값의 상관계수는 0.67로 나타났다.

### 3.7 위험도 등급화 및 공간조인

개발된 모델을 활용하여 사고 위험도를 점수화하였고, 위험도 데이터를 Jenks Natural Breaks Classification을 활용하여 지역별 위험도를 4가지 등급(위험, 경계, 주의, 관심)으로 분류하고 공간조인을 통해 <Figure 5>와 같이 시각화하였다. 그 결과 전체 119개 지역 중 위험등급이 10개 지역, 경계지역이 18개 지역, 주의등급이 31개 지역, 관심지역이 60개 지역으로 나타났고, 대부분의 위험지역이 온산지역과 와지지역에 분포하고 있었다. Jenks Natural Breaks Classification은 같은 등급 내 전체 값들의 평균을 기준으로 평균편차는 최소화되고, 각 등급 간의 분산은 극대화하는 것이다. 즉, 등급 내의 분산은 줄이고 등급 간의 분산을 최대화하는 방법이다[4].

위험등급의 지역 특성을 살펴보면, 총 업체 수는 61개 (산업단지 내 49%, 산업단지 외 51%)이고, 총 근로자 수는 4710명, 공장 노후도는 평균 13.6년(20년 이상 노후공장 21%)이었다.

사고에 영향을 끼치는 요인으로 노후도가 가장 큰 영향을 끼치고, 다음으로 자본, 지가, 종사자 수, 건물 높이 순으로 나타났다. 즉, 위험등급 지역은 공장이 노후화되고 영세업체가 많은 것으로 이해할 수 있다.



<Figure 5> Regional Risk Visualization

특히 국가산업단지인 온산국가산업단지의 특성을 살펴 보면 공장 노후도는 15.4년이며 그중 20년 이상 노후공장의 비중은 36%로 평균 이상이며 소규모 사업장의 경우 위험지역 내에서 65%를 차지하여 평균 이하로 대기업이 많이 밀집된 것으로 확인이 되었다. 이 지역에서는 소규모 사업장보다 노후 공장에 대한 집중 관리가 필요하다고 볼 수 있다.

와지산업단지의 위험지역의 특성을 살펴보면 공장 노후도는 13.5년이며 그중 20년 이상 된 노후공장의 비중은 23%로 평균적으로 보이며 소규모 사업장의 경우 위험지역 내에서 85%를 차지하고 있어 평균 이상으로 소규모 영세업체가 많이 밀집해 있는 것으로 보인다. 또한, 위험지역의 경우 공단 내부보다는 공단 주변에 많이 나타나고 있어 이 지역의 경우 공단 외부의 소규모 영세 사업장에 대한 집중 관리가 필요하다고 볼 수 있다.

## 4. 결론

본 연구에서는 고용노동부, 지자체를 비롯한 기관과 산업통상자원부(산업단지 공장정보), 건축물통합정보시스템 등 여러 곳에서 분산 관리되고 있는 산업재해, 공장정보 및 건물정보 등을 일정 공간단위로 융·복합하여 데이터베이스 구축하였다. 산업안전사고 및 공장 특성별 분포패턴을 시각화하고, 가장 많이 발생하는 산업단지 내 화재 사고에 대해서 사고 주요 요인 분석, 공간적 요인과 화재 사고와 상관관계 분석 및 앙상블기법 중 하나인 부스팅과 CART 분석이 결합한 분석 모델을 활용하여 지역별 위험도 도출하였다. 도출된 위험지수를 Jenks Natural Breaks Classification을 활용하여 “위험, 경계, 주의, 관심” 4가지 등급으로 분류하였다. 또한, 분석결과를 지도로 시각화하여 위험지역을 직관적으로 확인할 수 있다. 공간적 이질성, 공간적 의존성을 고려한 공간분석과 데이터마이닝의 융합 방법론을 통해 공간적 성질을 고려한 분석결과를 도출하였다.

본 연구의 한계는 산업재해데이터를 보유하고 있는 기관에서 관련 데이터 제공이 불가능하다는 방침에 따라 119 긴급출동데이터를 활용하여 분석하였다는 점이다. 향후 공공데이터 개방 및 활용이 활성화되면, 풍부한 데이터를 활용하여 화재사고뿐만 아니라 다양한 산업단지 내 사고에 대한 요인을 도출하고 사고유형별 사고위험도를 예측하는 추가 분석이 필요할 것이다.

## Acknowledgement

This research was supported by Basic Science Research

Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2017R1D1A1B03034475).

## References

- [1] Aselin, L., Spatial Data Analysis with GIS : An Introduction to App application In the Social Sciences, NCGIS, Technical Report 92-10, 1992, pp. 1-7.
- [2] Bae, G.H. and Yoo, H.H., Fire Occurrence Pattern Analysis and Fire Risk Calculation of Jinju City, *Journal of the Korean Society for Geospatial Information System*, 2014, Vol. 22, No. 4, pp. 151-157.
- [3] Cho, D.G. and Kim, Y.M., Analysis on Geographical Variations of the Prevalence of Hypertension Using Multi-year Data, *Journal of the Korean Geographical Society*, 2014, Vol. 49, No. 6, pp. 935-948.
- [4] Jenks, George F., The Data Model Concept in Statistical Mapping, *International Yearbook of Cartography*, 1967, Vol. 7, pp. 186-190.
- [5] Jo, D.G., GIS and Geographically Weighted Regression in the Survey Research of Small Areas, *Journal of the Korean Society of Surveying*, 2009, Vol. 10, No. 3, pp. 1-19.
- [6] Jo, S., Sung, H., and Ahn, B., A Comparative Study on the Performance of SVM and an Artificial Neural Network in Intrusion Detection, *Journal of the Korea Academia-Industrial Cooperation Society*, 2016, Vol. 17, No. 2, pp. 703-711.
- [7] Jung, H. and Kim, J., A Machine Learning Approach for Mechanical Motor Fault Diagnosis, *Journal of the Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 1, pp. 57-64.
- [8] Kang, T., Kim, J., Jang, J., and Hong, C., BIM-based Data Mining Model for Effective Energy Management, *Journal of Society of Korea Academia-Industrial Cooperation Society*, 2016, Vol. 16, No. 8, pp. 5591-5599.
- [9] Kim, H.Y. and Jun, C.M., Land Value Analysis Using Space Syntax and GWR, *Journal of the Korean Association of Geographic Information Studies*, 2012, Vol. 15, No. 2, pp. 35-45.
- [10] Kim, M.J., Optimal Selection of Classifier Ensemble Using Genetic Algorithms, *Journal of Intelligence and Information Systems*, 2010, Vol. 16, No. 4, pp. 99-112.
- [11] Kim, T.H., A Study of the Development and Utilization Plan of Volcanic Disaster Response System based on Spatial Information, *Journal of the Korea Academia-Industrial Cooperation Society*, 2014, Vol. 15, No. 12, pp. 7357-7363.
- [12] Kim, T. and Hong, S., Development of GIS based Air Pollution Information System using a Context Awareness Model, *Journal of the Korea Academia-Industrial Cooperation Society*, 2015, Vol. 16, No. 6, pp. 4228-4236.
- [13] Ko, K.S. and Yang, J.K., Industrial safety risk analysis using spatial analysis, *Spring symposium of Korea Industrial and System Engineering Society*, 2017, pp. 368-370.
- [14] Leem, Y.M., Hwang, Y.S., and Choi, Y.H., Factor Analysis on Injured People Using Data Mining Technique, *Journal of Korea Safety Management & Science*, 2005, Vol. 7, No. 4, pp. 61-71.
- [15] Lim, W., Kwon, K., Kim, J., Lee, J., and Cha, S., Comparison and Analysis of Anomaly Detection Methods for Detecting Data Exfiltration, *Journal of the Korea Academia-Industrial Cooperation Society*, 2016, Vol. 17, No. 9, pp. 440-446.
- [16] Ministry of Employment and Labor, Comprehensive plan for safety and health innovation in the industrial field, 2015, pp. 1-4.
- [17] Sim, G., Metadata Design Based on Vector Type Geospatial Information Standard for the Collection and Management of Inundation Map, *Journal of the Korea Academia-Industrial Cooperation Society*, 2016, Vol. 17, No. 5, pp. 42-48.
- [18] Yoo, E.H., The study of spatial statistical analysis in GIS environment, *Journal of Geography*, 1999, Vol. 34, pp. 25-47.

## ORCID

Kyeongseok Ko | <http://orcid.org/0000-0001-7598-185X>

Jaekyung Yang | <http://orcid.org/0000-0002-4904-1351>