

한국 제조업에서의 사업장별 산재패턴 군집분석을 통한 맞춤형 산재예방 네트워크 구축

김영선*·최성원*·조현정*·이경용*

1. Introduction

한국에서의 산업 재해현황을 보여줄 수 있는 지표 중 하나인 재해율은 1975년 공식 집계된 이래 4.39%였던 것이 매년 감소하여 2001년 0.77%, 2009년 0.70%를 나타내고 있다. 2009년 기준으로 13,884,927명의 근로자 중 97,821명이 산재가 발생하였으며 이중 2,181명이 사망하였다. 특히 제조업은 가장 많은 재해자가 발생시키는 산업으로써 2009년 기준으로 32,997명의 재해자가 발생하여 전체 재해의 약 34%를 차지하고 있다.

한국은 산업구조의 급속한 변화에 따라 고용과 생산의 측면에서의 주력산업이 제조업이었으나 탈공업화(deindustrialization) 현상을 통해 서비스업 중심의 경제구조로 개편되고 있다. 서비스업이 중심산업이 됨에 따라 재해예방활동 역시 많은 변화가 필요하다. 따라서 효율적 재해 예방활동을 위해서는 재해패턴이 동질성이 있는 그룹을 찾아서 이들의 속성을 연구하고 이를 기반으로 산재예방정책을 마련해야 한다. 한국에서 기존 재해패턴의 동질성을 나타내는 기준으로는 사업장이 속해 있는 업종과 사업장의 규모를 기준으로 대상사업장을 분류하여 산재예방 활동을 하였다. 하지만 이러한 기준이 실제 사업장의 외면적 특성에서는 동질성을 나타내고 있지만 산재 특성에는 동질성을 나타내지는 못하는 실정이다. 따라서 산재 특성의 동질성을 나타내는 기준을 찾아내는 것이 중요하다. 하지만 동질성을 설명해줄 수 있는 다양한 정보를 수집하고 모형화하기 위해서는 굉장히 많은 정보를 수집해야 하기 때문에 시간과 비용적 측면에서 효율성이 떨어지게 된다. 따라서 본 연구에서는 과거 재해율의 패턴이 동질성을 갖도록 군집화하여 이들 집단의 특성을 연구하고 재해예방의 기초자료로서의 활용가치에 대해 분석하고자 한다.

* 한국산업안전보건공단 산업안전보건연구원

2. Method

2.1 Data Collection

한국산업안전보건공단에서 산재 예방활동의 기초자료로 사용하기 위해 운영중인 사업장 DB과 재해자 DB는 산업재해보상보험법에 적용받는 사업장과 재해자에 대한 다양한 정보를 가지고 있다. 사업장 DB에는 id에 해당하는 관리번호와 사업장 일반현황에 해당하는 업종, 규모, 근로자수, 지역 정보, 사업장의 산재현황에 해당하는 재해자수, 사고부상자수, 질병이환자수, 사고사망자수, 질병 사망자수의 정보를 포함하고 있다. 2001년 산재보험에 가입된 사업장의 수는 10,581,186개이며 약 9년 사이에 약31%의 사업장이 산재보험에 가입되어2009년에는 13,884,927개로 집계되고 있다. 이중 제조업 사업장의 수는 2009년 기준 257,686개소로써 전체 사업장의 16%를 차지하고 있다.

본 연구에서는 300인 이상의 제조업 사업장을 대상으로 하였으며 2001년에서 2009년 사이의 재해율을 목적변수로 하였다. 사업장의 규모를 300인 이상으로 한 이유는 한국에서의 재해율이 약 0.7%를 나타내고 있기 때문에 사업장별 재해율 변동이 있기 위해서는 사업장의 규모가 일정 수준 이상이어야 한다. 데이터 전처리 과정을 통해 9년간 폐업하지 않거나 동일 id인 관리번호를 유지하는 사업장의 수는 295이다.

표1. 연도별 사업장과 근로자, 재해자 수의 분포

연도	전규모			제조업			300인이상 제조업		
	근로자수	사업장수	재해자 수	근로자수	사업장 수	재해자 수	근로자 수	사업장 수	재해자 수
2001	10,581,186	909,461	81,434	2,923,112	187,326	35,525	846,633	806	4,812
2002	10,571,279	1,002,263	81,911	2,858,337	194,219	34,941	772,543	768	4,964
2003	10,599,345	1,006,549	94,924	2,840,359	198,492	40,230	717,384	757	7,305
2004	10,473,090	1,039,208	88,874	2,929,998	214,002	37,599	728,552	756	7,517
2005	11,059,193	1,130,094	85,411	3,054,064	222,841	36,033	769,150	759	5,481
2006	11,688,797	1,292,696	89,910	3,033,228	236,492	35,946	683,257	693	4,278
2007	12,528,879	1,429,885	90,147	3,096,006	247,522	34,151	676,866	657	3,851
2008	13,489,986	1,594,793	95,806	3,104,554	255,135	35,848	633,838	638	3,850
2009	13,884,927	1,560,949	97,821	3,182,262	257,686	32,997	668,599	624	2,922

2.2 Data Analysis Model

2.2.1 재해율 패턴 Fitting과 1단계 Clustering

300인 이상 제조업 295개의 사업장의 9년간 재해율 패턴을 살펴보면 모두 제각기 다른 형상을 나타내고 있다. 따라서 이들 형상들간의 최소의 공통점을 찾기 위해서는 재해율의 대체적 추세를 찾아내야 한다. 다양한 회귀분석을 통해 개별 사업장의 재해율 y 은 모형 추정치 \hat{y} 과 모형에 의해 설명하지 못하는error term ϵ 으로 구성된다. 따

라서 ϵ 를 최소화 하는 모형을 찾아내야 하는데 이를 위해 linear fn.과quadratic fn., exponential fn., logarithm fn., sine fn.으로 그림1과 같이 형태를 다양화하였다. 295개의 사업장을 5가지 함수에 의해 fitting한 후 모형에 의한 변동량과 우연변동량간의 차이에 대한 F 통계량을 최대화하는 함수를 선택하였다.

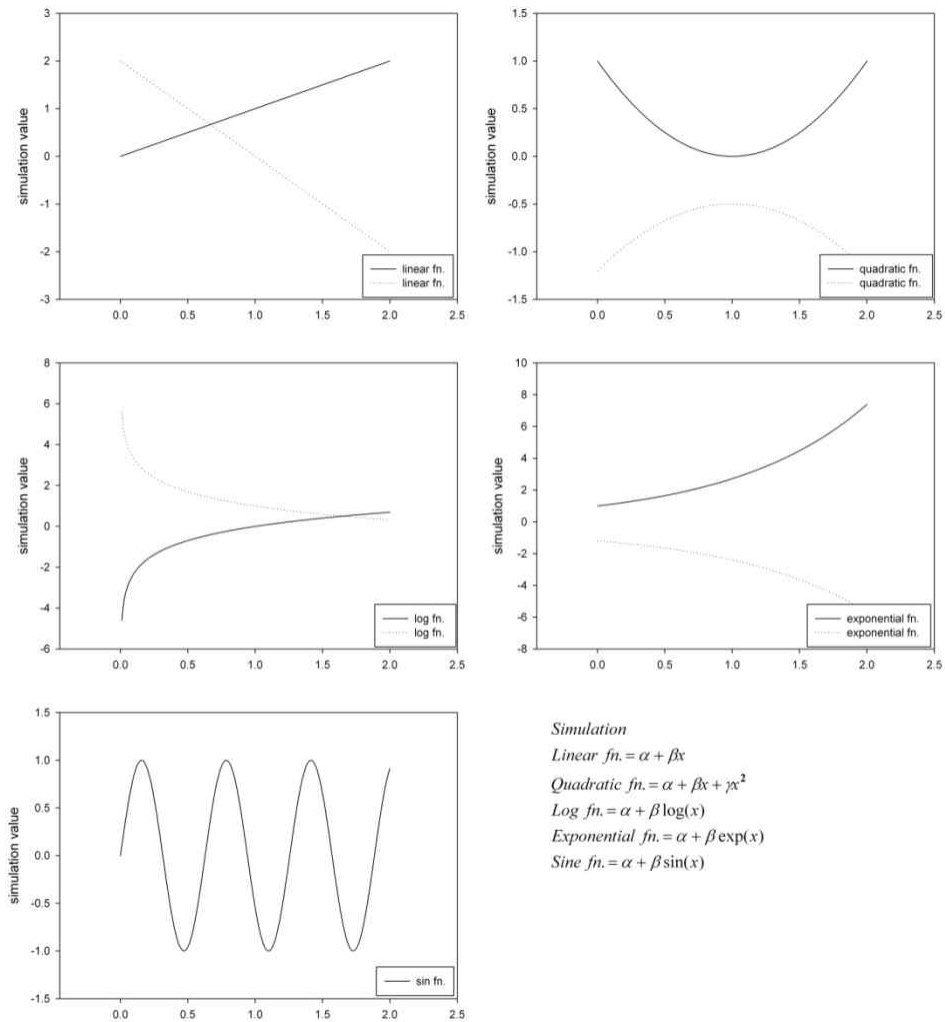


그림 1. 5가지 함수에 의한 재해율 패턴 시뮬레이션

2.2.2 재해율 패턴에 대한 계수의 변환

재해율 패턴을 5개의 함수로 1차 분류한 후 각 Cluster내에 있는 사업장들의 재해율 패턴을 2차 군집분석하였다. Linear fn.으로 분류된 사업장의 경우 α, β 에 따라서 다양

하게 분류해 볼 수 있다. α 의 크기에 따라 과거 재해율이 높은 사업장인지 낮은 사업장인지를 추정해 볼 수 있다. β 값이 양수인 경우에는 해당 사업장의 재해율이 증가 추세를 나타내고 있으며 음수인 경우에는 감소 추세를 나타내고 있다. Quadratic fn.으로 분류된 사업장의 경우에 γ 가 양수인 경우에는 재해율이 감소후 증가의 추세를 나타내고 있으며 음수인 경우에는 증가후 감소의 추세를 나타내는 사업장이다. Log fn.의 경우에는 β 값이 양수인 경우 재해율 증가의 추세를 나타내고 있지만 최근들어 증가율이 둔화되어진 사업장이며 β 값이 음수인 경우는 재해율이 감소의 추세를 나타내고 있지만 최근들어 감소의 감소율이 둔화된 사업장들이다. Exponential fn의 경우에는 β 값이 양수인 경우에는 재해율이 과거 둔화된 상태에서 증가하였으나 최근들어 증가율이 급증한 사업장이며 β 값이 음수인 경우에는 과거 둔화된 상태에서 재해율이 감소하였으나 최근들어 증가율이 급증한 사업장이다. 이와 같이 5개의 함수에 회귀모형에 의한 fitting 한 후 각coefficient인 α, β, γ 에 따라서 재해율이 서로 다르게 나타나게 된다. 따라서 보다 동질성 있는 재해율 패턴을 추정하기 위해 5개의 함수별로 α, β, γ 를 변수로 하여 군집분석을 실시하고자 한다. 하지만 재해율의 증가와 감소를 나타내는 coefficient인 β 의 경우 양수인지 음수인지 여부에 따라서 앞서 해석한 바와 같이 매우 다른 의미를 부여하게 된다. 하지만 기존의 군집분석시 알고리즘을 살펴보면 0을 중심으로 β 값의 절대치가 매우 작은 값들에 대해서는 동일 Cluster로 묶여지게 된다. 따라서 이러한 문제점을 먼저 해결하기 위해 β 를 sigmoidal fn.을 통해 식(1)와 같이 변환시켰다. Sigmoid fn.은 S자 형태로 값들을 변환시켜 특정 값 주변에 밀집되어져 있는 값의 경우는 그 차이를 극대화 시키고 일정 수준 이상의 값들의 차이는 완화시켜 주는 역할을 한다.

$$x' = \frac{a}{1 + e^{-\frac{x-c}{b}}} + d \dots\dots\dots(1)$$

a는 sigmoid function의 범위, b는 shape, c와 d는 위치와 관련된 coefficient이며 295개 사업장의 β coefficient를 살펴본 결과 a = 2, b=0, c=0.01, d= -1를 선택하여 사용할 경우 군집분석시 발생하는 문제점을 해결할 수 있었다.

2.2.3 군집분석

재해율 패턴 fitting 분석을 통해 추정된 parameter coefficient에 의해 만들어진 p차원 공간에 흩어진 N개의 사업장에 대해 어떤 의미의 조밀성을 가지고 cluster를 이루고 있는지를 분석하고자 한다. 군집분석의 방법은 군집의 형태와 사용되는 similarity의 척도와 연관되어 다양한 방법이 있다. Similarity와 non-similarity와 관련하여 개체들 간의 거리를 활용하여 측정하게 되는데 거리가 가까우면 상사성이 큰 것을 의미한다. 이때 사용되는 거리로써는 Euclidean distance, Minkowski distance, Mahalanobis

distance, Canberra distance, czekanowski distance 등이 있다. 이 중 Euclidean distance, Minkowski distance와 같은 경우는 척도불변성(scale invariance)를 가지고 있지 않기 때문에 분석시 주의를 해야 한다. 이에 대한 문제점의 해결안으로 각 변수를 해당 표준편차로 나누어 측정단위를 없애는 표준화를 생각할 수도 있으며 일반화한 측도로써 변수들의 상관관계도 고려한 Mahalanobis distance를 식(2)와 같이 사용할 수 있다.

$$D_{MH} = \|z_r - z_s\| = (x_r - x_s)'S^{-1}(x_r - x_s) \dots \dots \dots (2)$$

Mahalanobis distance에 의해 얻은 거리행렬 D를 가지고 각 사업장들을 가까운 개체들끼리의 모임인 Cluster를 만들어내야 하는데 이러한 과정은 병합과 분할의 과정으로 크게 나뉘어지며 알려진 방법으로는 최단연결법, 최장연결법, 중심연결법, 중위수연결법, 평균연결법, Ward 방법등으로 나뉘어진다. 최단연결법의 경우에는 계산시간이 비교적 짧고 고립된 군집을 찾는데 우수하며 최장연결법은 군집들의 응집성에 중점을 두고 있다. ward방법은 개체들을 하나의 군집으로 묶음으로써 생기는 정보의 손실을 최소화할 수 있는 방법이다. 따라서 본 연구에서는 비교적 적은 개체(사업장)을 분석하기 때문에 ward의 방법을 선택하였다.

3. Result

3.1 대규모 제조업 사업장에 대한 재해율 분석

295개 300인이상 제조업 사업장에 대해 재해율을 계산한 결과 2003년까지 증가의 추세를 나타내다가 이후 매년 감소하여 2009년 0.49%를 나타내고 있다. 제조업 전체를 대상으로 2009년 재해율의 경우는 1.04%와는 많은 차이를 나타내고 있다. 이는 제조업에서의 산재가 소규모사업장을 중심으로 많이 발생하는 것을 의미한다. 반면 개별기업을 단위로 하여 근로자수를 가중치로 재해율을 계산할 경우 발생하는 재해율에 대한 편차를 살펴보면 2003년 68%, 2008년 62% 등으로 매우 높은 값을 나타내고 있다. 이는 재해발생이 사업장들 간에 차이가 매우 큰 것으로써 재해예방이 잘 된 기업과 그렇지 못한 기업들이 모두 혼재하고 있음을 의미한다.

3.2 재해율 패턴 clustering

9년간 사업장의 재해율을 5개의 함수를 이용하여 회귀분석을 통해 1단계 Clustering을 한 결과 linear fn.으로 분류된 사업장의 수는 29개이고 quadratic fn.으로 분류된 사업장의 수는 50개, log fn.으로 분류된 사업장의 수는 95개, exponential fn.으로 분류된 사업장의 수는 62개, sine함수로 분류된 사업장의 수는 54개로 계산되었다. 또한 9

년간 재해가 거의 발생하지 않은 사업장은 5개로써 이들 사업장에 대해서는 따로 분류하였다. 이들 분류된 사업장의 회귀계수들을 sigmoidal fn.으로 변환한 결과 그림 2와 같이 나타나고 있다.

표2. 규모300인 이상 제조업에서의 재해율 분포

연도	사업장 수	근로자 수	재해자 수	재해율	재해편차
2001	295	435,247	2,781	0.64%	28%
2002	295	448,474	3,538	0.79%	37%
2003	295	428,930	4,906	1.14%	68%
2004	295	444,409	5,338	1.20%	66%
2005	295	455,322	3,866	0.85%	44%
2006	295	436,312	3,055	0.70%	36%
2007	295	442,902	2,764	0.62%	32%
2008	295	409,975	2,888	0.70%	62%
2009	295	440,986	2,154	0.49%	29%

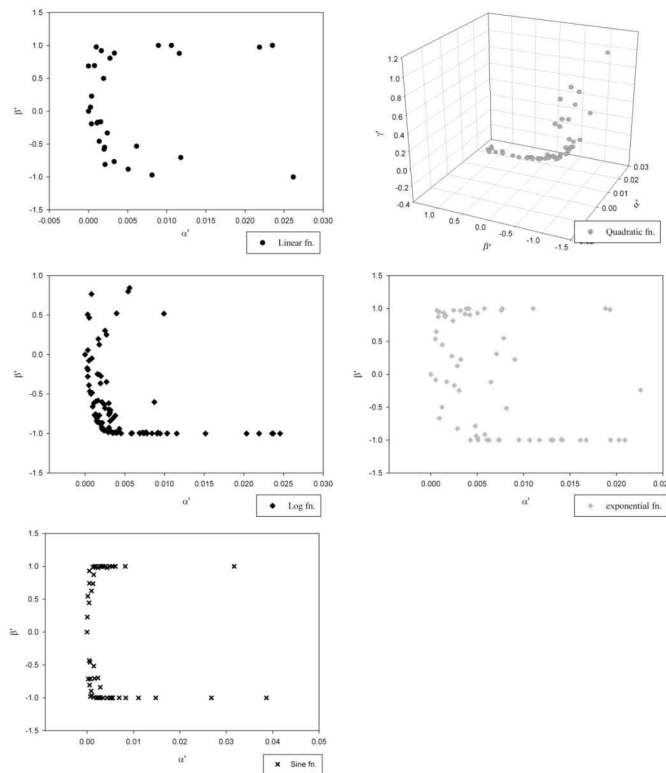


그림2. 재해율 패턴에 의한 1차 clustering 후 회귀계수들의 분포도

재해율의 패턴에 따라 1단계에서 분류된 사업장을 대상으로 하여 5개의 fn.별로 군집분석을 한 결과 그림3과 같은 결과가 나타난다. 각 cluster에 속한 사업장 295개를 각각 9년간 재해율 패턴을 그려보고 이들의 특징을 살펴보았다. Linear fn.의 경우에는 3개의 cluster들이 만들어지는데 1번 cluster의 경우에는 재해율 증가를 하고 있으며 2번 cluster의 경우는 감소, 3번 cluster는 재해율이 매우 낮은 상태에서 감소를 나타내고 있다. Quadratic fn.의 경우 1번 cluster의 경우 감소 후 증가의 패턴을 나타내고 있었으며 2번 cluster의 경우에는 증가 후 감소, 3번 cluster의 경우에는 증가 후 감소를 하되 증가와 감소의 폭이 매우 크게 나타나고 있었다. Log fn.의 경우에는 1번 cluster는 재해율이 매우 낮은 상태에서 감소의 추세를 나타내고 있었고 2번 cluster의 경우는 재해율이 높은 상태에서 감소의 추세를 나타내고 있고 3번 cluster는 재해율이 낮은 상태에서 증가의 추세를 나타내고 있었다. Exponential fn.의 경우에는 1, 2번 cluster의 경우는 감소를 3번 cluster는 증가의 추세를 나타내고 있다. Sine fn.의 경우에는 매년 변화의 폭이 상당히 심하게 나타나고 있어 실제 그 추세를 파악하기 어려운 것으로 나타나고 있다.

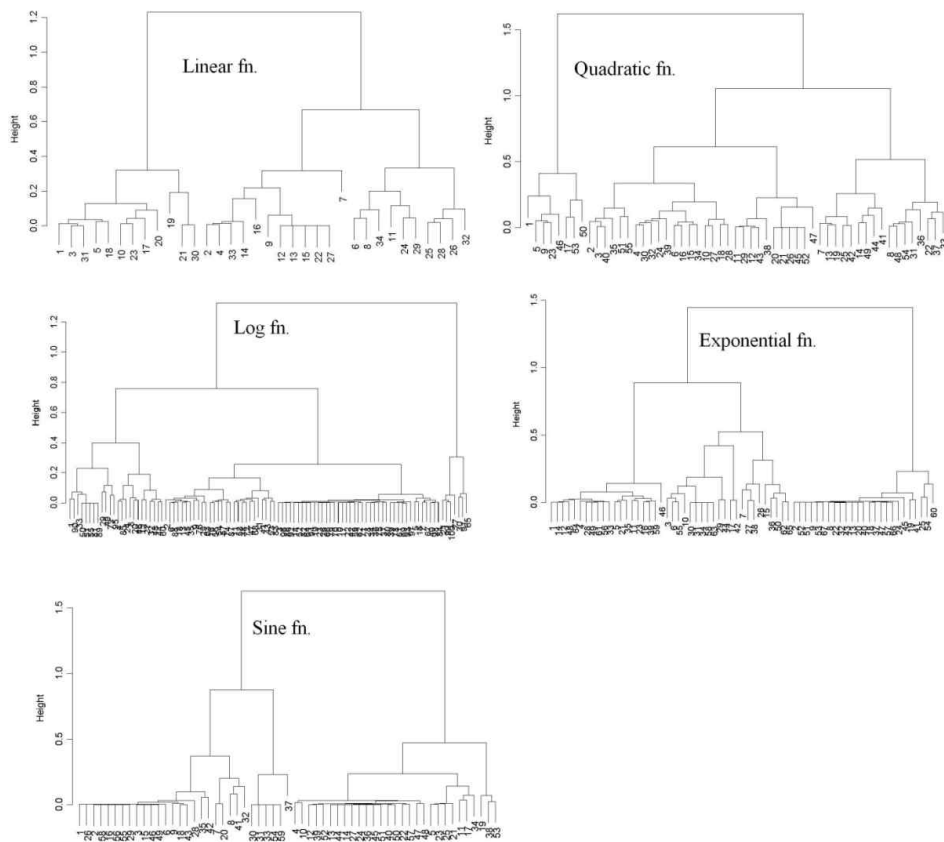


그림3. 재해율 패턴에 대한 2차 clustering 분석 결과

3.3 Cluster를 이용한 정보교류 네트워크 구축

2단계 clustering 분석을 통해 재해율의 패턴별 동질성이 있는 14개의 cluster를 생성하였다. 이들을 재해감소율과 재해율간의 그림을 그려보면 그림4와 같이 나타난다. 먼저 재해가 발생하지 않거나 아주 낮은 수준으로 발생하는 사업장을 재해예방의 Leader로 볼 수 있다. 이러한 사업장들은 기업내 재해예방의 솔루션과 문화조성이 잘 구축되었기 때문에 다른 cluster에 속해 있는 사업장들이 벤치마킹해야 하는 대상이다.

또한 사업장간 네트워크를 형성을 통해 서로간 정보 공유를 통해 산재를 예방해야 할 것이다. 사업장 혹은 기업은 영리를 추구하는 이익단체이며 유사분야의 타 사업장 혹은 기업은 경쟁대상일수 있다. 하지만 재해예방은 근로자의 생명을 대상으로 하기 때문에 경쟁구도 보다는 서로간 상생의 구도를 만들어야 할 것이다. Linear(3) 는 비교적 낮은 수준의 재해가 발생하고 있으며 과거에 비해 지속적으로 재해율이 떨어지는 사업장이다. 이러한 사업장 역시 재해가 다소 높은 수준에서 감소(Linear(2))하고 있거나 과거 재해가 과거에 비해 많은 수준으로 낮게 발생하고 있지만 재해율이 다소 높은 사업장(Exponential(1))에 솔루션 공유를 통해 서로간 정보공유를 하는 것이 좋을 것이다. Quadratic(2,3)과 같은 경우에는 과거 재해율이 증가하였지만 일정 시점을 기준으로 감소를 하는 사업장이다. 이러한 사업장의 경우는 재해증가 요인을 잘 분석하여 회사에 맞는 재해예방 솔루션을 구축하였고 이를 통해 지속적으로 감소를 시키고 있는 사업장이다. 따라서 과거 재해가 감소하였지만 최근들어 지속적으로 증가(Quadratic(1))하고 있거나 재해율의 추이를 예측할 수 없을 정도로 fluctuation이 심하게 나타나는 사업장(Sine))에게 과거 재해율 추세를 변화시킬 수 있었던 요인에 대해서 서로간 정보 공유가 필요할 것이다. 이와 같이 재해율 패턴을 통해 동질성 높은 Cluster와 이들의 특징을 통해 기업간 재해예방 네트워크를 형성시켜 볼 수 있다.

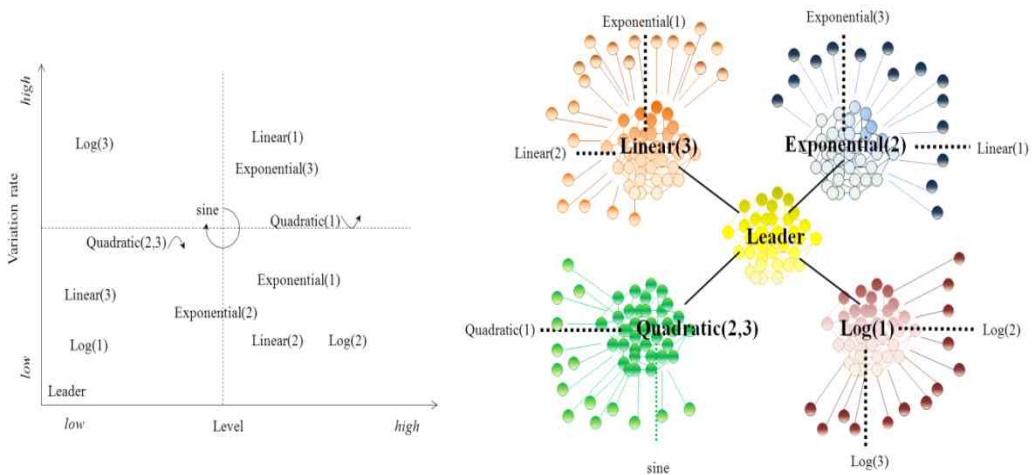


그림4. 재해율 수준과 증감율에 따른 Cluster의 분포와 정보교류 네트워크 구성

4. Discussion

재해의 동질성이 있는 사업장을 구성하기 위해 방법으로써 과거 사업장이 소속된 업종과 규모를 기준으로 하여 cluster를 구성하였지만 이들 cluster내의 사업장들의 재해율은 서로 상이하였다. 이러한 원인으로써 사업장의 재해를 나타내는 속성이 단지 업종과 규모 외에도 기업의 정책적 측면과 안전보건관리자의 활성화 정도, 근로자의 교육과 실행적 측면, 사업장내 위험인자에 대한 솔루션 등 아주 많은 요인들이 관여하고 있다. 이를 조사하고 계량화하여 동질성 있는 cluster를 구성할 수 있지만 이는 시간과 비용이라는 측면을 고려할 경우 현실적인 대안일 수는 없다. 본 연구에서는 이러한 다양한 요인들이 사업장내 재해율이라는 지표로써 결과물을 산출한다는 가정아래 재해율의 패턴을 통해 동질성 있는 사업장의 cluster를 구성하였다. 또한 동질성 있는 cluster를 기업간 네트워크로 구성하여 서로간 솔루션 공유를 통해 재해예방을 위한 정책적 제안을 본 연구에는 포함되어 있다. 앞서 언급한 바와 같이 기업은 이익의 극대화를 목표로 하고 있으나 재해라는 측면은 기업의 본질적 속성으로 해결해서는 안된다. 특히 근로자의 생명을 다루고 있는 분야에 대해서는 공익적 측면으로 접근해야 하며 이를 통해 근로자의 삶과 기업의 안정적 노동력 공급, 국가의 발전에 기여될 것이다.