Journal of Korea Spatial Information Society Vol.20, No.6 : 23-32, December 2012 ISSN 2287-9242(Print)
http://dx.doi.org/10.12672/ksis.2012.20.6.023 ISSN 2287-9250(Online)

자기조직화지도와 GIS를 이용한 다차원 공중보건자료의 탐구적 분석

Exploring Multidimensional Public Health Data Using Self Organizing Map and GIS

손 철*
Chul Sohn

요 약 본 연구에서는 2005년에서 2010년까지의 통계청 남성암 유형별 연령표준화 시군구 사망률 데이터에 대해 자기조직화지도와 GIS를 이용한 탐구적 자료 분석을 수행하여 이들 데이터에 의미 있는 패턴이 내재되어 있는지 분석하였다. 그리고 지역의 사회경제적 수준을 대표하는 변수로 선정된 지역별 가구주의 교육수준과 분석된 패턴이 어떤 관련이 있는지 검토하였다. 분석결과 우리나라 시군구는 남성암 사망원인 측면에서 독특한 특성을 가진 18개의 지역 군집으로 구분될 수 있으며, 이들 군집 내 속한 시군구가 공간적으로도 군집되는 경향이 있음을 보여주었다. 또한 가구주의 교육수준이 높은 군집이 낮은 군집에 비해 남성암 사망률이 낮은 경향을 보이지만 일부 암의 경우 교육수준이 높은 군집에서 사망률이 높음을 보여주었다. 이 결과는 지역의 사회경제적 요인, 자연환경적 요인 등 암의 발생 및 관리에 영향을 미치는 지역적 요인에 양의 공간적 자기상관이 존재하며, 이러한 공간적 자기상관이 다양한 유형의 암 원인 사망에 영향을 미친 결과로 해석되어질 수 있다. 또한 18개의 군집 중 서울의 강남구 및 서초구를 포함한 군집은 대부분 유형의 암 원인 사망률에서 전체 18개 군집 중 하위 수준임을 보여 우리나라 암의 예방, 발생, 관리와 관련된 중요원인이 사회경제적 요인일 수 있음을 암시하였다.

키워드 : 남성암 사망률, 자기조직화 지도, 지리정보시스템

Abstract This study applied an exploratory analysis based on Self Organizing Map and GIS to cause specific age-standardized regional death rates data related to ten types of male cancers to find meaning patterns in the data. Then the patterns revealed from the exploratory analysis was evaluated to investigate possible relationship between these patterns and regional socio-economic status represented by regional educational attainment levels of head of household. The results from this analysis show that SI-GUN-GUs in Korea can be clustered to eighteen unique clusters in the stand point of male cancer death rates and these clusters are also spatially clustered. Also, the results reveal that regions with higher socio-economic status show lower level of the death rates compared with the regions with lower socio-economic status. However, for some cancer types, the regions with higher socio-economic status show relatively higher death rates. These patterns imply that the prevention, detection, and treatment of male cancers might be strongly affected by regional factors such as socio-economic status, environmental factors, and cultures and norms in Korea. Especially, one of the eighteen clusters, which includes Gangnam-Gu and Seocho-Gu, shows lower death rates in many of male cancer types. This implies that socio-economic status may be one of the most influential factors for regional cancer control.

Keywords : Male Cancer Death Rates, Self Organizing Map, GIS

* Associate Professor, Dept. of Urban Planning and Real Estate, Gangneung-Wonju National University, csohn@gwnu.ac.kr(This paper is the modified version of the paper presented in KSIS conference, Spring, 2012..)

1. 서 론

지역의 인구구성, 연령, 교육정도, 고용수준, 자연환경, 의료서비스 수준, 문화 및 규범 등 다양한 요인이 지역사회의 건강수준에 영향을 미친다[5]. 그리고 지역사회의 건강수준은 지역보건통계에 반영된다. 따라서 지역보건통계에 대한 분석을 통해 지역 간 건강수준의 격차에 대한 분석이 가능할 뿐 아니라 이 격차를 가져오는 다양한 지역적 요인에 대한 추론이 가능하다.

우리나라 통계청 사망원인 통계, 질병관리본부 지역사회건강조사, 건강보험공단 지역의료통계 등은 시군구별로 세분화된 다차원의 공중보건관련 데이터를 제공한다. 이 데이터들은 세분화된 지역의 공중보건수준에 대한 다양한 추론을 하는 데 이용될 수 있다.

탐구적 데이터 분석(Exploratory Data Analysis)은 연구자에게 데이터의 본질적 특성은 잘 보전하면서 데이터에 내재해 있는 구조나 패턴을 드러내주는 연구방법론이다[7].

본 연구에서는 2005년에서 2010년까지의 통계청 남성암 유형별 연령표준화 시군구 사망률 데이터에 대해 자기조직화 지도와 GIS를 이용한 탐구적 자료분석을 수행하여 지역별 남성암 유형별 사망률 데이터에 내재된 의미 있는 패턴을 찾을 수 있는지 분석하였다.

2. 자기조직화 지도와 GIS를 이용한 탐구적 분석

2.1 자기조직화 지도[1)]

자기조직화 지도(Self Organizing Map: SOM)는 다차원의 데이터를 연구자가 이해하기 쉬운 형태로 시각화하고 이를 통해 연구자가 데이터 내에 숨겨진 의미 있는 패턴을 찾아낼 수 있도록 하는 환경을 제공한다[3].

기본적으로 자기조직화 지도는 n차원의 투입 벡터(input vector)를 2차원 공간에 맵핑시킨다. 2차원의 공간은 보통 6각형 셀을 기본단위로 한 그리드 형태로 구성된다. 각 셀은 셀을 대표하는 값인 가중치 벡터(weight vector)를 갖는다. 자기조직화 지도는 n 차원의 투입 벡터를 이용하여 2차원 공간상의 가중치 벡터를 훈련(train)한다. 훈련이 종료되면 2차원 공간상에서 근접한 셀들은 유사한 가중치 벡터 값을 가지게 된다. 이러한 특성이 지도라는 명칭을 자기조직화 지도가 가지게 되는 이유이다. 그리고 훈련된 가중치 벡터는 투입 벡터를 대표하는 값을 가진다.

2차원 공간의 훈련이 종료되면 추가적으로 U-Matrix와 Component Plane이란 결과물을 얻을 수 있다. U-Matrix는 가중치 벡터를 이용하여 이웃한 셀 간의 거리를 계산하여 2차원 공간상에 표시한 것이다. U-Matrix를 통해 투입 벡터를 몇 개의 군집으로 구분하는 것이 타당한가에 대한 정보를 얻을 수 있다. Component Plane은 2차원 공간상에 가중치 벡터를 구성하는 변수의 값을 표시한 것이다. 투입 벡터가 n 개의 변수를 가질 때 n 개의 Component Plane이 존재한다. 변수상호간에 긴밀한 상관관계가 존재할 경우 Component Plane의 패턴도 유사한 패턴을 보이게 된다. 즉 n 차원의 투입 벡터에 대해서는 n 개의 Component Plane이 존재하므로 이것들을 상호 비교하면 변수들 간의 상관관계를 파악할 수 있다.

2차원 공간의 훈련이 종료되고 가중치 벡터가 최종적으로 완성되면 훈련에 사용된 투입 벡터는 가장 유사한 값을 가진 셀에 소속시킬 수 있다. 이렇게 해서 투입 벡터의 1차적 군집이 형성된다. 그런데 2차원 공간을 구성하는 셀이 많으면 형성되는 군집의 수도 많게 된다. 따라서 군집의 수를 좀 더 줄이기 위해 셀을 대표하는 가중치 벡터에 K-means Clustering 등의 방법을 적용하여 몇 개의 셀을 포함하는 2차적 군집을 형성하는 방법이 일반적으로 사용된다. 그리고 2차적 군집의 형성 시에는 다양한 군집의 개수 가운데 Davies-Bouldin index (v_{DB})[2)]를 최소화하는 군집의 개수를 최종적인 군집의 수로 선택하는 방법을 일반적으로 이용한다[19].

1) Agarwal and Skupin[1], Lu et al.[13], Sohn[17] 등에 근거하여 설명이 이루어졌다.

2) $v_{DB}(C) = \frac{1}{C}\sum_{K=1}^{C}\max_{l \neq k}\left\{\frac{S_c(Q_k)+S_c(Q_l)}{d_{ce}(Q_k, Q_l)}\right\}$,
C: 군집의 수
$S_c = \sum_i \| X_i - c_k \| / N_k$: 군집중심과 군집에 포함된 데이터간의 거리
c_k: 군집 k의 중심값
N_k: 군집 k에 존재하는 데이터의 수
$d_{ce} = \| c_k - c_i \|$: 군집 Q_k와 Q_l의 거리

자기조직화 지도에서 가중치 벡터의 훈련은 다음과 같은 단계에 의해 이루어진다[13].

1단계: 가중치 벡터($\theta_j(0)$)를 임의로 초기화 한다. t=0

단, t: 반복시간

2단계: 1개의 투입 벡터와 모든 가중치 벡터와의 유클리드 거리(d_j)를 계산한다.

$d_j = \| x_i - \theta_j(t) \|$

3단계: 유클리드 거리가 가장 적은 Best Matching Unit(c)을 찾는다.

$u_c = \min\{d_j\}$

4단계: 모든 가중치 벡터의 값을 다음 식에 의해 수정한다.

$\theta_j(t+1) = \theta_j(t) + \eta(t) h_{cj}(t) [x(t) - \theta_j(t)]$

단, $h_{cj}(t) = \exp\left(-\frac{\| r_j - r_c \|^2}{2\sigma^2(t)}\right)$

:이웃함수(시간에 따라 감소)

$r = \| r_j - r_c \|$: j셀과 BMU c의 속성 거리

$\eta(t)$: 학습률(시간에 따라 감소)

$\sigma(t)$: 이웃 범위

5단계: 시간을 t에서 t+1로 증가시키다. t<T일 경우 2단계로 간다.

단, T: 미리 정해진 최대 반복수

2.2 자기조직화 지도와 GIS의 결합

기본적으로 자기조직화 지도를 통해서 개별 투입 벡터가 어떤 군집에 소속되는지, 소속된 군집이 어떤 특성을 가지는지, 변수들 간에 어떤 상관관계가 있는지 등의 질문에 답할 수 있다.

자기조직화 지도로부터 얻어진 결과를 GIS의 시각화 기능과 공간분석 기능을 통해 분석함으로써 다음과 같은 추가적 질문에 답할 수 있다[18]. 첫째, 동일한 군집에 속하는 투입 벡터가 지리 공간적으로 집적하는가? 둘째, 지리 공간적으로 이웃한 투입 벡터가 동일한 군집에 속하는가?(유사한 속성을 가지는가?) Figure 1은 자기조직화 지도와 GIS가 결합함으로써 구체적으로 어떤 연구 질문들에 답할 수 있는지 보여준다.

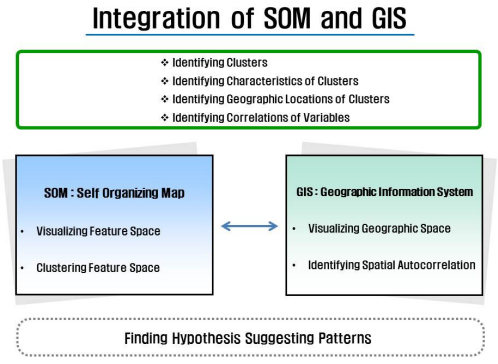

Figure 2. Self Organizing Map and GIS

3. 관련연구에 대한 검토

최근 우리나라의 지역보건통계 데이터를 분석한 대표적 연구에는 Choi et al.[4], Sim and Son[16], Kim et al.[8], Jeong et al.[6], Kim and Park[9]의 연구가 있다. 이들 연구들에서는 건강수준의 지역적 격차를 가져오는 요인이 무엇인가에 대한 분석이 주로 이루어졌다.

Choi et al.[4]은 부산시 동별, 원인별 표준화 사망률 (Standard Mortality Ratio)과 동별 사회적 결핍지수와의 관계를 Gaussian Conditional Autoregressive (CAR)모델을 이용하여 분석하였다. 분석결과는 사회적 결핍지수(Deprivation Index)[3]가 높은 곳(사회경제적 수준이 낮은 곳)에서 심장질환과 부상으로 인한 사망률이 높음을 보였다. 그러나 동 연구에서는 암 사망률과 사회적 결핍지수(Social Deprivation Index)와의 관련성을 찾지 못했다. 동연구에서는 암 사망률과 사회적 박탈지수와의 관계를 발견하지 못한 주요한 이유로 유형별 암 사망률에 대한 분석이 이루어지지 못한 것을 들고 있다.

Sim and Son[16]은 강원도 시군구 자료를 이용한 분석에서 뇌혈관계 표준화 사망률과 물질적 결핍지수, 병의원의 수, 도로포장률, 지방세 수준과의 상관관계를 분석하였으나 통계적으로 유의한 관계를 발견하지 못하였다.

Kim et al.[8]은 2001년 서울시 보건지표 자료를 이용하여 동별 파워엘리트의 입지계수와 개인의 흡

3) 지역의 사회경제적 수준에 대한 지수

연률과의 관계를 분석하였다. 동 연구에서는 사회경제적 지위가 낮은 곳에서 흡연률이 높음을 발견하였다.

Jeong et al.[6]은 96개 시군에 대해 지역의 물질적 결핍지수와 15세에서 64세 인구의 표준화 사망비의 상관관계를 1995년에서 2000년 사이 자료를 이용하여 분석한 결과 물질적 결핍수준과 사망비 사이에 정의 상관관계가 존재함을 발견하였다.

Kim and Park[9]은 1990년, 1995년, 2000년 시군별 센서스 자료를 이용한 분석에서 시군의 조사망률, 표준화 사망률에 교육수준과 보건수준이 부의 영향을 미친 것을 보고하였다.

이상의 연구들은 사회경제적 요인이 현재 우리나라의 지역적 보건수준 격차의 주요한 원인임을 시사해주고 있지만 다음과 같은 한계를 가진다. 첫째, 분석에 이용된 사망원인이나 질병정보가 몇 가지에 한정되며, 동일유형의 사망원인에서도 지나치게 집계된 데이터를 이용하여 지역의 상황에 따라 다양한 질병이 어떻게 발생하는가에 대한 다차원적 분석이 이루어지지 않았다. 즉, 다양하게 생산되는 지역보건 통계정보가 연구에서 충분히 이용되지 못했다.[4] 둘째, Kim and Park[9]을 제외하면 대부분의 연구가 부산, 강원도, 서울, 전국의 일부지역 등 일부지역의 분석에 한정하고 있다. 따라서 우리나라 전체지역에 대한 결론을 이끌어내지 못하고 있다. 셋째, 대부분의 연구가 한 해의 보건통계만을 분석에 이용하여 장기적 추세에 대한 분석이 이루어지지 않았다.

본 연구는 선행연구와 달리 남성암에 대한 보다 세분화된 분류기준을 이용하고 장기적 추세가 반영된 전국 시군구 자료를 대상으로 분석을 수행하여 앞선 연구들의 한계를 보완하고자 한다.

4. 데이터 및 분석설계

4.1 데이터

본 연구에서 사용된 데이터는 통계청에서 발간하는 사망원인통계로부터 획득되었다. 통계청 사망원인통계는 우리 국민의 사망원인구조를 파악하기 위한 목적으로 작성되며 사망신고서 자료를 기초로 한다. 본 연구는 통계청의 사망원인통계 중 시군구별 10개 유형의 남성암 연령표준화 사망률[5] 데이터를 분석의 기초 자료로 이용하였다. 그리고 분석 시에는 지역별 사망률의 장기적 추세를 반영하기 위해 시군구별 2005년도에서 2010년 데이터의 6개년 평균을 구하여 이용하였다. Table 1은 본 연구에서 고려된 남성암 유형을 보여준다.

Table 1. Types of Male Cancers

Code	Definition
C15	Malignant neoplasm of oesophagus
C16	Malignant neoplasm of stomach
C18_C21	Malignant neoplasms of colon, rectum and anus
C22	Malignant neoplasms of liver and intrahepatic bile ducts
C25	Malignant Neoplasm of pancreas
C33_C34	Malignant neoplasms of trachea, bronchus and lung
C50	Malignant neoplasm of breast
C61	Malignant neoplasm of prostate
C70_C72	Malignant neoplasms of other part of central nervous system
C91_C95	Leukaemia

Table 2는 분석에서 사용된 남성암 유형별, 6개년 평균 246개 지역 사망률의 통계량을 보여준다.

Table 2. Statistics for Average Six Year Cancer Death Rates

Variable	Mean	Std. Dev.	Min	Max
c15	6.348171	2.4061	1.333333	19.21667
c16	31.92209	5.388941	18.36667	51.78333
c18_c21	16.71321	2.934352	8.516667	25.6
c22	36.85888	8.087556	22.93333	60.45
c25	9.158401	1.849152	3.233333	15.38333
c33_c34	50.73747	7.88741	30.06667	74.43333
c50	0.062669	0.135806	0	0.8
c61	5.634417	1.499387	1.783333	13.38333
c70_c72	2.688144	1.037978	0.35	7.683333
c91_c95	3.743767	1.22057	0	9.383333

4.2 분석설계

자기조직화 지도와 GIS의 기능을 이용하여 우리

4) Choi et al. [4]에서 암의 유형별 사망률을 고려하지 못하고 전체 암 사망률만을 사용한 것이 대표적인 사례이다.

5) $\frac{\Sigma(\text{연령별 사망률} \times \text{표준인구의 연령별 인구})}{\text{표준인구}} \times 100{,}000$

나라 지역별 남성암 사망률 데이터에 내재한 패턴을 파악하기 위해 본 연구에서는 Figure 2와 같은 분석방법을 사용하였다.

먼저 10개 남성암 사망률 연령표준화 지표를 이용하여 자기조직화 지도를 1차적으로 훈련하고, 여기서 얻어진 U-Matrix와 Component Plane을 이용하여 10개 남성암 유형별 사망률 사이에 주목할 만한 상관관계가 존재하는가를 분석하였다.

다음으로 최종적으로 훈련된 자기조직화 지도의 가중치 벡터와 시군구별 사망률 데이터를 이용하여 Best Matching Unit을 파악하였다. 이를 통해 유사한 유형별 암 사망률 패턴을 공유하는 시군구를 파악하였다. 또한 가중치 벡터에 K-Means 클러스터링 알고리즘을 적용하여 2차 군집화를 시도하였다. 2차 군집화 수행시 최적군집의 수를 선택하기 위해 최대 군집의 수를 20개로 설정하고 Davies-Bouldin index(v_{DB})를 이용하는 방법을 사용하였다.

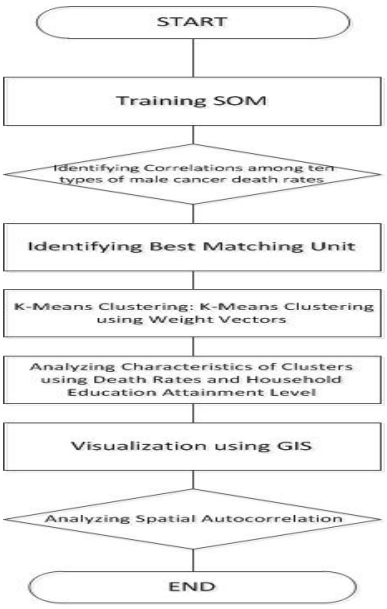

Figure 3. Flow of Analysis

식별된 군집들에 대해서는 각 군집에 속하는 시군구가 10개 유형의 남성암 사망률 측면에서 어떤 평균적 특징을 가지고 있는가를 분석하였다. 또한 각 군집에 속한 시군구가 어떤 사회경제적 특성을 가지고 있는가를 파악하기 위해서 군집에 속하는 시군구들의 대학 이상 교육수준의 가구주 평균 비율(UNIV) 및 대학원 이상 교육수준의 가구주 평균 비율(POST)을 이용하였다.[6] 본 연구에서 지역의 사회경제적 특성을 대표하는 변수로 가구주의 교육수준을 선택한 것은 교육수준이 건강수준에 영향을 미치는 소득과 밀접한 연관이 있을 뿐만 아니라 건강한 삶을 살아가는 데 필요한 시간사용과 소비에 필요한 지식을 공급하는 역할을 하기 때문이다[2]. 해석의 편의를 위해 각 군집의 평균 암 유형별 사망률과 가구주 교육수준 비율은 시군구 암유형별 사망률과 가구주 교육수준의 백분위수를 기준으로 H(75백분위수 이상), MH(75-50백분위수), ML(50-25 백분위수), L(25 백분위수 이하) 등 사단계로 구분되어 해석에 이용되었다.

마지막으로 자기조직화 지도를 통해 형성된 군집을 GIS를 이용하여 지도상에 맵핑하여 어떤 공간적 패턴이 존재하는지를 시각화와 통계적 방법을 통해 검증하였다. 공간적 군집에 대한 통계적 검증을 위해서는 Moran's I가 이용되었다. Moran's I 분석을 위해서 동일군집에 속한 시군구의 속성으로 1을 부여하였고 속하지 않은 시군구의 경우 0을 부여하였다. 이웃의 정의를 위해서는 Polygon Contiguity 기준을 이용하였다.

5. 분석결과

5.1 U-matrix 와 Component Plane을 이용한 분석

통계청에서 시군구별로 제공하는 11개 암 관련 연령표준화 사망률 자료 중 남성에게서는 발생하지 않는 자궁암을 제외한 나머지 10개 암 유형에 대한 246개 시군구 자료를 이용하여 자기조직화 지도가 훈련되었다.[7] 훈련결과 얻어진 최종적 자기조직화 지도의 U-Matrix와 Component Plane은 Figure 3와 같다. Figure 3의 자기조직화 지도는 84(12×7)개의 육각형 셀로 구성되었으며 각 셀을 구성하는 가중치 값을 가지고 있다.

가중치 벡터의 차원은 투입 벡터의 차원과 같은

6) 2010년 센서스 결과를 기준으로 하였다.

7) Matlab SOM Toolbox (URL: http://www.cis.hut.fi/projects/somtoolbox/)를 훈련에 이용하였다. SOM Toolbox의 명령어인 som_make()를 이용하여 자기조직화지도의 차원이 결정되었다. 훈련을 위해 사용된 투입자료는 246×10 차원의 지역별 6개년 평균 남성암 유형별 사망률 매트릭스이다. 10개 변수는 훈련전 0과 1사이 값으로 변환되었다.

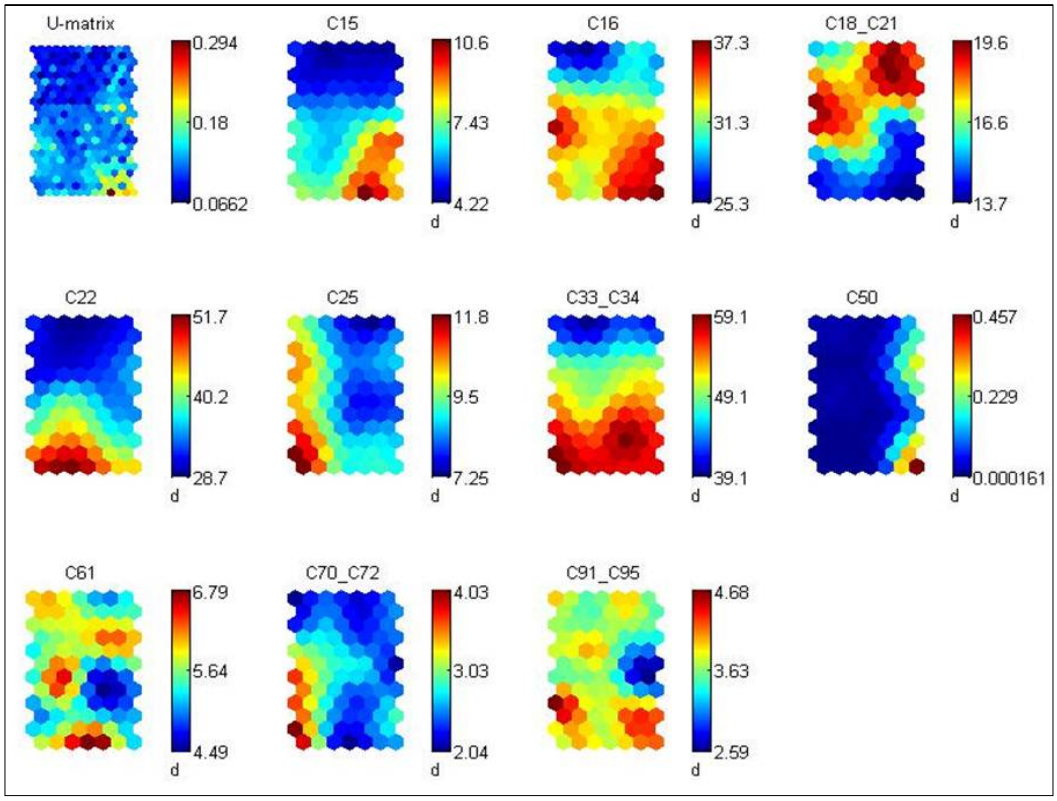

Figure 3. U-Matrix and Component Plane

10차원이다. Figure 3의 10개 Component Plane은 84개 셀 상에 가중치의 벡터를 구성하는 변수의 값들이 어떻게 분포해 있는지를 보여준다. U-Matrix를 살펴보면 자기조직화 지도의 우하단에 인접 셀과 속성에 다른 셀들이 분포함을 보여준다. 10개의 Component Plane에서 동일한 지역을 살펴보면 이 지역은 C15, C16, C33_C34, C50, C91_95의 연령표준화 사망률이 높으며 C18_C21, C25, C70_72의 연령표준화 사망률이 낮은 지역임을 알 수 있다.

Figure 3의 10개 Component Plane의 시각화 패턴을 비교할 때 남성암 유형별 사망률 간에 주목할 만한 상관관계는 찾을 수 없으나 C15 와 C16, C22와 C33_C34 사이에 유사한 패턴이 존재할 가능성이 있음을 파악할 수 있다.

5.2 군집화를 통한 분석

Figure 4는 Figure 3의 84개 셀을 K-Means 클러스터링 방법을 이용하여 CL1에서 CL18의 18개의 대분류 군집으로 다시 2차 분류한 결과를 보여준다.[8] 자기조직화 지도에서는 유사한 속성을 가진 셀이 공간적으로도 인접해 있으므로 18개 군집도 모두 공간적으로 인접한 셀로만 구성되어 있음을 보여준다.

CL15 군집에 속하는 셀 25(BMU)에는 용인시 수지구, 서울시 강남구, 서초구, 송파구, 성남시 분당구 등 5개 지역이 속한다. 이것은 사회경제적 수준이 유사한 이들 지역에서 남성암 유형별 사망률 패턴 역시 유사함을 의미한다. 반면 송파구와 접해 있는 강동구는 CL3 군집의 셀 38(BMU)에 속한다. 셀 38(BMU)에는 서울시 중구가 소속되어 강동구와 중구의 남성암 사망률 패턴이 유사함을 시사한다.

5.3 GIS를 통한 공간분석

Figure 5에서 Figure 7은 자기조직화 지도에서

8) K-Means 클러스터링을 위해서는 SOM Toolbox의 kmeans_clusters()의 기능이 이용되었다. 최대 가능한 군집수는 20개로 설정되었고 2개에서 20개의 군집수별로 각각 500회의 K-Means 클러스터링을 반복하여 이 가운데 분류오류를 최소화하는 군집을 해당군집수의 최적군집으로 선정하고 선정된 19개의 군집 가운데 v_{DB}가 최소인 18개 군집을 최종적 군집수로 선택하였다.

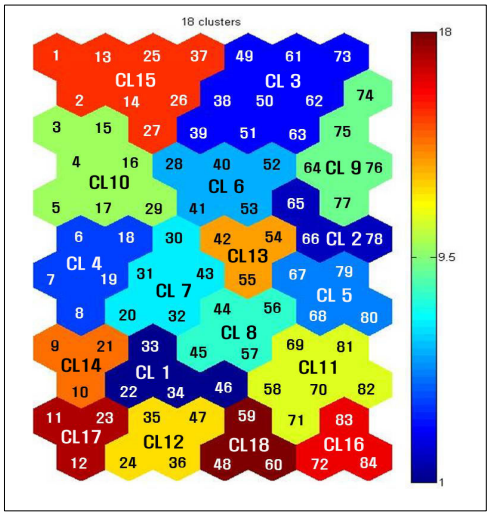

Figure 5. Result of Re-Classification

얻어진 18개 군집별로 각 군집에 소속된 시군구를 지도상에 나타낸 것이다. 동 지도는 우리나라 시군구별 가구 중 대학원 과정 이상 교육수준의 가구주 비율도 보여준다. Figure 6는 CL15 군집에 속하는 시군구가 주로 고학력 가구주 비율이 높은 지역이며 공간적으로도 서울 강남권과 인접지역에 군집해 있음을 보여준다. Figure 7은 CL7 군집의 경우 부산지역에 공간적으로 군집화 되어 있음을 보여준다. CL7 군집은 Table 3에서 볼 수 있는 바와 같이 지역 교육수준이 중상이며 암 사망률도 보통수준인 지역이다.

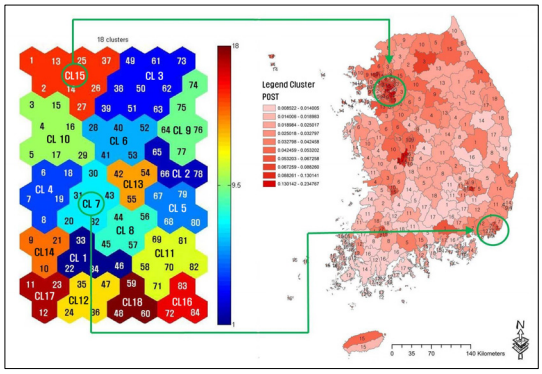

Figure 6. Spatial Distribution of CL7 and CL15 Clusters

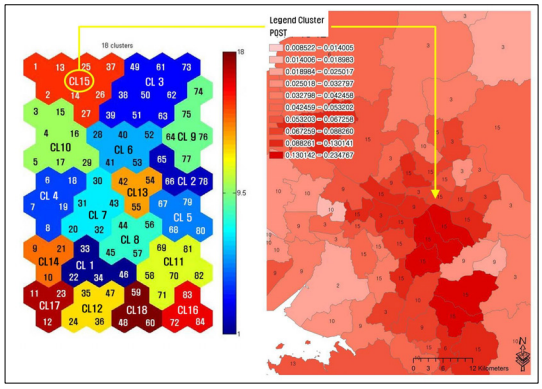

Figure 7. Spatial Distribution of CL15 Cluster

Table 3. Characteristics of Clusters

Cancer / Cluster	c00_c97	c15	c16	c18_c21	c22	c25	c33_c34	c50	c61	c70_c72	c91_c95	UNIV	POST
CL1	MH	MH	ML	ML	H	MH	MH	L	ML	MH	ML	ML	ML
CL2	MH	MH	MH	ML	MH	L	MH	H	ML	L	L	ML	ML
CL3	L	L	ML	H	L	L	L	MH	ML	ML	MH	MH	H
CL4	H	MH	H	H	MH	H	MH	MH	ML	H	ML	ML	ML
CL5	ML	H	MH	L	ML	L	MH	MH	L	MH	L	ML	ML
CL6	ML	ML	ML	MH	ML	ML	MH	MH	MH	ML	MH	MH	MH
CL7	MH	MH	MH	MH	H	MH	ML	MH	H	MH	ML	MH	MH
CL8	MH	MH	MH	MH	MH	L	H	L	L	L	MH	ML	ML
CL9	ML	ML	ML	H	MH	ML	ML	H	MH	MH	ML	MH	H
CL10	ML	ML	ML	H	ML	H	ML	MH	ML	MH	MH	MH	MH
CL11	MH	H	H	ML	MH	ML	H	H	ML	MH	H	L	ML
CL12	H	MH	ML	ML	H	MH	H	MH	H	ML	MH	ML	ML
CL13	ML	MH	H	ML	MH	L	ML	L	L	H	ML	ML	ML
CL14	MH	MH	MH	L	MH	H	H	L	ML	H	H	ML	ML
CL15	L	L	L	ML	L	ML	L	MH	MH	ML	ML	H	H
CL16	H	H	H	L	H	MH	H	H	ML	ML	MH	L	L
CL17	H	H	H	ML	H	H	H	L	ML	H	ML	L	L
CL18	H	H	H	L	H	ML	H	L	H	L	MH	L	ML

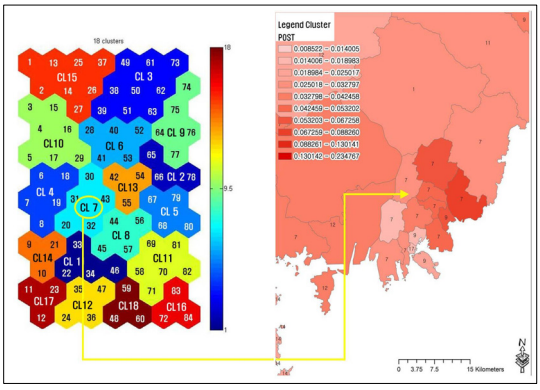

Figure 8. Spatial Distribution of CL7 Cluster

Figure 5에서 Figure 7은 우리나라의 경우 남성암 원인 사망률이 유사한 지역이 공간적으로도 집적하는 패턴을 보이는 것을 보여준다.

Table 4. Results from Moran's I

cluster	Moran's I Index	Z score
CL1	0.04	1.19
CL2	-0.03	-0.66
CL3	0.16	4.02*
CL4	0.03	0.93
CL5	0.03	0.9
CL6	0.01	0.4
CL7	0.53	13.23*
CL8	0.02	0.59
CL9	0.03	0.72
CL10	0.12	2.96*
CL11	0.14	3.54*
CL12	0.14	3.69*
CL13	0.02	0.54
CL14	0.01	0.27
CL15	0.28	6.9*
CL16	0.08	2.03*
CL17	0.03	0.76
CL18	0.13	3.51*

*:Significant at the 5 percent level of significance (Island regions are excluded for Moran's I Analysis)

이러한 공간적 군집현상은 각 군집에 대한 Moran's I 분석결과를 통해서도 잘 드러난다. 18개 군집에 대한 Moran's I 추정결과를 정리한 Table 4는 8개 군집에서 5% 신뢰수준에서 통계적으로 유의한 정의 공간적 자기상관이 존재함을 보여준다.

5.4 분석결과의 종합

본 연구의 분석결과는 Table 5와 같이 종합적으로 정리될 수 있다. Table 5는 자기조직화 지도로부터 얻어진 CL1에서 CL18 군집별로 10개 유형의 6개년 평균 암 사망률이 H, MH, ML, L로 분류된 개수와 해당 군집의 전체 암 사망률(C00_C97), 대학이상 학력 가장비율, 대학원이상 학력 가장비율을 H, MH, ML, L로 분류한 내용, 그리고 각 군집별 Moran's I 분석결과를 보여준다.

Table 5로부터 확인할 수 있는 점은 4가지이다. 첫째, 가장의 교육수준이 높은 지역의 경우 대부분의 유형별 암 사망률의 수준이 낮다는 점이다(CL3와 CL15의 경우). 둘째, 가장의 교육수준이 낮은 지역의 경우 유형별 암 사망률이 높다는 점이다(CL16과 CL17의 경우). 셋째, 6개년 남성암 사망률의 비공간적 속성으로만 형성된 군집들이 지리 공간적으로도 군집하는 경향이 있다는 점이다. 특히 이러한 공간적 군집경향은 학력이 높은 군집에서 두드러지게 나타났다(CL7과 CL15의 경우). 넷째, 가장의 학력이 높은 군집에서도 일부 유형 암의 경우 높은 사망률을 보이며(CL3의 경우) 학력이 낮은 군집에서도 일부 유형의 암의 경우 사망률이 낮은 것으로 나타났다.

Table 5. Summarized Results

Cluster	c00_c97	H	MH	ML	L	univ	post	Moran's I Index
CL1	MH	1	4	4	1	ML	ML	0.04
CL2	MH	1	4	2	3	ML	ML	-0.03
CL3	L	1	2	3	4	MH	H	0.16*
CL4	H	4	4	2	0	ML	ML	0.03
CL5	ML	1	4	1	4	ML	ML	0.03
CL6	ML	0	5	5	0	MH	MH	0.01
CL7	MH	2	6	2	0	MH	MH	0.53*
CL8	MH	1	5	0	4	ML	ML	0.02
CL9	ML	2	3	5	0	MH	H	0.03
CL10	ML	2	3	5	0	MH	MH	0.12*
CL11	MH	5	2	3	0	L	ML	0.14*
CL12	H	3	4	3	0	ML	ML	0.14*
CL13	ML	2	2	3	3	ML	ML	0.02
CL14	MH	4	3	1	2	ML	ML	0.01
CL15	L	0	2	4	4	H	H	0.28*
CL16	H	5	2	2	1	L	L	0.08*
CL17	H	6	0	3	1	L	L	0.03
CL18	H	5	1	1	3	L	ML	0.13*

*:Significant at the 5 percent level of significance

6. 결 론

본 연구에서는 2005년에서 2010년까지의 통계청 남성암 유형별 연령표준화 시군구 사망률 데이터에 대해 자기조직화 지도와 GIS를 이용한 탐구적 자료분석을 수행하여 이들 데이터에 의미 있는 패턴이

내재되어 있는지 분석하였다.

분석결과 우리나라 시군구는 남성암 사망 원인 측면에서 독특한 특성을 가진 18개의 지역 군집으로 구분될 수 있으며, 이들 군집 내 속한 시군구가 지리 공간적으로도 군집되는 경향이 있음을 보여주었다. 이 결과는 지역의 사회경제적 요인, 자연환경적 요인 등 암의 발생 및 관리에 영향을 미치는 지역적 요인에 양의 공간적 자기상관이 존재하며, 이러한 공간적 자기상관이 다양한 유형의 암 원인 사망에 영향을 미친 결과로 해석되어질 수 있다.

또한 18개의 군집 중 가구주의 학력으로 대표되는 사회경제적 수준이 높은 지역의 경우 대부분의 유형의 암 원인 사망률에서 전제 18개 군집 중 하위 수준임을 보였다. 이 점은 우리나라 암의 예방, 발생, 관리와 관련된 중요원인이 사회경제적 요인일 수 있음을 암시하는 것이다. 그러나 특정 암의 경우 오히려 이 지역에서 높은 사망률을 보여 높은 사회경제적 수준이 반드시 모든 유형의 암을 억제하지 않을 가능성도 보여주고 있다.

탐구적 분석의 목적은 가설을 제시하고 검정하기보다 검정될 새로운 가설을 제안하는 데 있다[3]. 본 연구에서 드러난 지역별 남성암 사망률 데이터의 패턴은 남성암의 발생, 발견, 관리에 영향을 미치는 다양한 지역적 요인을 통계적으로 확인하는 후속 연구들에 기초자료로 이용될 수 있을 것이다.

References

[1] Agarwal, Pragya; Skupin Andre 2008, Self Organizing Maps: Applications in Geographic Information Science, Ed. West Sussex, England: Wiley.

[2] Arendt, Jacob Nielsen. 2005, Dose Education Cause Better Health, Economics of Education Review 24:149-160.

[3] Arribas-Bel, Daniel; Nijkamp, Peter; Scholten, Henk 2011. “Multidimensional Urban Sprawl in Europe: A Self-Organizing Map Approach,” Computers, Environment and Urban Systems 35:263-275.

[4] Choi, Min-Hyeok; Cheong, Kyu-Seok; Cho, Byung-Mann; Hwang, In-Kyung; Kim, Chang-Hun; Kim, Myoung-Hee;p Hwang, Seung-Sik; Lim, Jeong-Hun; Yoon, Tae-Ho 2011, Deprivation and Mortality at the Town Level in Busan, Korea: An Ecological Study, Journal of Preventive Medicine and Public Health, 44(6):242-248.

[5] Flowerdew, Robin; Manley, David J.; Sabel, Clive E. 2008, Neighbourhood Effects on Health: Does it matter where you draw the boundaries?, Social Science & Medicine 66:1241-1255.

[6] Jeong, Baek-Geun; Jung, Kap-Yeol; Kim, Joon-Youn; Moon, Ok-Ryun; Lee, Yong-Hwan ; Hong, Young-Seoub; Yoon, Tae-Ho 2006, The Relationship between Regional Material Deprivation and the Standardized Mortality Ratio of the Community Residents Aged 15-64 in Korea, Journal of Preventive Medicine and Public Health, 39(1):46-52.

[7] Kaski, S.; Kohonen, T. 1996, Exploratoy Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World,” Proceeding of the Third International Conference on Neural Network in the Capital Markets, London, England, October 11-13, 1995.

[8] Kim, Chang-Seok; Yun, Sung-Cheol; Kim, Hye-Ryun; Kim, Young-Ho 2006, A Multilevel Study on the Relationship between the Residential Distribution of High Class (Power Elites) and Smoking in Seoul, Journal of Preventive Medicine and Public Health, 39(1): 30-38.

[9] Kim, Du-Seop; Park, Hyo-Jun 2003, Regional Differentials in Mortality in Korea, 1990-2000, The Population Association of Korea, 26(1): 1-30.

[10] Kim, Kang-Ho 2009, The Effect of Schooling and Training Participation on Wage of Employed Workers, Journal of Agricultural Education and Human Resource Development, 41(3):123-151.

[11] Kim, Yoo-Sun 2009, Wage Determinants in the Korean Labour Market, Korean Journal of

Industrial Relations, 19(2):1-25.

[12] Kim, Young-Chang; Chang, Jae-Woo 2009, Countinuous k-Nearest Neighbor Query Processing Algorithm for Distributed Grid Scheme, Journal of Korea Spatial Information System Society, 11(3):9-18.

[13] Lu; Hsin-Chung; Chang, Chung-Liang; Hsieh, Jen-Chieh 2006, Classification of PM_{10} distributions in Taiwan, Atmospheric Environment 40:1452-1463.

[14] Park, Ji-Man; Hwang, Chul-Sue 2005, A Design and Practical Use of Spatial Data Warehouse for Spatial Decision Making, The Journal of GIS Association of Korea, 13(3):239-252.

[15] Shin, Kyung-Soo; Choi, Chang-Yeoul 2007, A Study on the Wage Gap by the Education Standards and Gender Productivity Gap, Productivity Review, 21(3):97-121.

[16] Sim, Jeoung-Ha; Son, Mi-A 2009, The Relationship between Cerebrovascular Mortality and Community Health Indicators in Gangwon-do, Journal of Agricultural Medicine and Community health, 34(1):1-12.

[17] Sohn, Chul 2011, Identification of Spatial Housing Submarkets in the Capital Region Using Self Organizing Map and Chow Test, The Korea Spatial Planning Review, 70:151-166.

[18] Spielman, S. E.; Thill, Jean-Claude 2008, Social Area Analysis, Data Mining, and GIS, Computers, Environment and Urban Systems 32:110-122.

[19] Vesanto, Juha; Alhoniemi, Esa 2000, Clustering of the Self-Organizing Map, IEEE Transactions on Neural Networks, 11(3):586-600.

논문접수 : 2012.06.11
수 정 일 : 1차 2012.09.11 / 2차 2012.12.05
심사완료 : 2012.12.06

Chul Sohn

1992 Master in City Planning, Seoul National University
2000 Ph. D. in Urban and Regional Science, Texas A&M University
2002 Engineer, Samsung SDS Public Development Division
2002~Present Associate Professor, Dept. of Urban Planning and Real Estate,, Gangneung-Wonju National University
Research Expertise
• Spatial Analysis and Data Mining