Journal of the Korean Data & Information Science Society
2013, **24**(4), 783–791
http://dx.doi.org/10.7465/jkdi.2013.24.4.783
한국데이터정보과학회지

서포트벡터 군집분석을 이용한 대구·경북지역 대학의 분류

박혜정[1] · 김종태[2]

[1]대구대학교 공통교양 · [2]대구대학교 전산통계학과

접수 2013년 5월 13일, 수정 2013년 6월 14일, 게재확정 2013년 7월 1일

요 약

본 논문에서는 대구·경북지역의 24개 4년제 대학교의 대학공시센터에 공시한 대학지표 자료를 사용하였다. 이들 대학지표들 중 재학생 충원률과 건강보험 취업률에 대한 지표를 이용하여 유사 특징을 가지고 있는 대학들을 그룹화 (분류)하여 그룹의 특징을 분석하는데 목적이 있다. SPSS의 계층적 군집분석과 서포트벡터 군집분석을 분석방법으로 활용하여 실험한 결과에서 공통으로 도출할 수 있는 정보를 구하였다.

주요용어: 가우시안 커널, 군집분석, 대학지표, 서포트벡터 군집분석.

1. 서론

통계적 학습이론 (statistical learning theory)에 기반을 둔 서포트벡터기계 (support vector machines; SVM)에 의한 분석방법은 주어진 문제를 항상 전역적 최적의 해를 구하기 때문에 패턴인식 분야에서 활용도가 높다. 또한 분류 효율적 측면에서는 로지스틱 회귀나 판별분석의 확률추정에 기반을 둔 방법들 보다 서포트벡터기계에 의한 분석방법이 전반적으로 예측력이 좋다는 것은 잘 알려진 사실이다. 서포트벡터기계 관련 연구는 Shim와 Seok (2013), Hwang과 Shim (2012), Hwang (2011)에 잘 나타나 있다.

서포트벡터기계의 기본 원리는 선형 분리 (linearly separable)가 가능한 문제에서부터 출발 하는데 이는 다시 크게 서포트벡터 회귀 (함수)추정과 서포트벡터 군집분석으로 나누어진다. (Vapnik, 1995, 1998). 서포트벡터 군집분석 알고리즘은 데이터 점들을 가우시안커널 (Gaussian kernel)을 이용해 데이터 공간에서 고차원 특징공간 (feature space)으로 사상 (mapping)한다. 서포트벡터 군집분석 알고리즘은 특징공간에서 데이터들의 상 (image)을 둘러싸고 있는 가장 작은 구 (sphere)를 찾으며, 이 구는 데이터 공간으로 사상되어지고 데이터 점들을 둘러싸고 있는 등고선 (contour)의 집합을 형성한다. 이 등고선들은 군집 경계 (boundary)들로 해석되며, 각 분리된 등고선에 둘러싸여져 있는 점들은 같은 군집를 갖게 된다. 경계 기반의 군집을 분류하는 방법들을 확장하여 분석하는 기법들이 많이 연구되었다. 서포트벡터 군집분석 (support vector cluster; SVC) 기법으로는 Yang 등 (2002)에 제안한 SVC 기반의 proximity graph 방식인 DD (delaunay diagram), MST (minimum spanning tree), k-NN (k-nearest neighbors)와 Lee와 Lee (2005)에서 제안한 SVC 기반의 dynamical system인 SEP-CG (stable equilibrium point-complete graph) 방식이 있다.

[1] (712-714) 경북 경산시 진량읍 대구대로 201, 대구대학교 공통교양, 초빙교수.
[2] 교신저자: (712-714) 경북 경산시 진량읍 대구대로 201, 대구대학교, 전산통계학과, 교수.
E-mail: jtkim@daegu.ac.kr

본 논문에서는 대구·경북지역의 4년제 대학 24개교의 공시되어 있는 대학지표 자료를 가지고 유사 특징을 가지고 있는 대학들을 그룹화 (분류)하여 그룹의 특징을 분석하고자 한다. 일반적으로 군집분석을 하기 위해서는 SPSS Statistics 18 버전의 계층적 군집분석을 많이 활용하고 있다. 본 논문에서는 SPSS Statistics 18 버전의 계층적 군집분석과 함수 추정에 로버스트하다고 알려진 서포트벡터 군집분석을 사용하였다. 서포트벡터 군집분석으로는 Yang 등 (2002)에 제안한 proximity graph 방식인 DD, MST, K-NN과 Lee와 Lee (2005)에서 제안한 dynamical system인 SEP-CG 방식을 사용하였다.

본 연구의 목적은 한국대학교육협의회 대학정보공시센터의 '대학알리미'에서 중요 대학지표들에 대하여 유사 특징을 가지고 있는 대학들을 그룹화(분류)하기 위하여 SPSS의 계층적 군집분석과 서포트벡터 군집분석 방법들을 활용하여 공통으로 도출할 수 있는 정보를 분석하고. 그룹의 특징을 살펴보고자 한다.

본 논문은 다음과 같이 구성하였다. 2절에서는 서포트벡트 군집분석에 대한 기본원리를 기술하였으며, 3절에서는 한국대학교육협의회 대학정보공시센터의 대학 알리미에서 공시되어 있는 대구·경북지역의 4년제 대학 24개교의 대학지표 자료를 가지고 대학별 군집분석을 하였다. 마지막으로 4절에서는 실험을 통한 결론을 기술하였다.

2. 서포트벡터 군집분석

서포트벡터 군집분석 알고리즘은 Schölkopf 등 (2001)과 Tax와 Duin (1999)에 의해 데이터 집합의 서포트벡터 설명을 공식화 하였다. d-차원 입력 공간에 존재하는 개체 $x \subseteq R^d$로 구성되는 n개의 데이터 공간의 자료 집합을 $\{\mathrm{x}_i\} \subseteq x$라고 하자. 비선형 변환 (nonlinerar transformation) Φ를 사용해서 x를 고차원 특징 공간으로 사상하여 반경 R로 가장 작게 둘러싸고 있는 구를 찾는다.

$$\|\Phi(x_i) - \alpha\|^2 \leq R^2 \quad \forall i.$$

단, $\|\cdot\|$는 유클리디언 놈 (Euclidean norm)이고 , α는 구의 중심이다. ξ_i는 j-번째 입력 데이터 x_i가 구에서 벗어나는 정도를 정량적으로 나타내는 양이라 할 때 위의 식은 다음과 같이 표현할 수 있다.

$$\|\Phi(x_i) - \alpha\|^2 \leq R^2 + \xi_j.$$

여기서 $\xi_j \geq 0$이다. 이 문제를 풀기위해 라그랑즈 (Lgrangian) 함수를 사용한다.

$$L = R^2 - \sum_j (R^2 + \xi_j - \|\Phi(x_j) - \alpha\|^2)\beta_j - \sum \xi_j \mu_j + C \sum \xi_j.$$

여기에서 $\beta_j(\geq 0)$와 $\mu_j(\geq 0)$은 라그랑즈 승수 (Largrangian multiplier)이며, C는 벌칙 상수이다. 최적화 문제는 구의 반경 R^2과 입력 자료가 구를 벗어난 경우에 부과되는 총 벌점 관련항 $\sum \xi_j$의 가중치 합 (weighted sum)으로 구성되어 있다. 벌칙 상수 C는 R^2과 $\sum \xi_j$의 상대적 중요성을 조정하는 상수 (trade-off constant)이다. 최적 해는 라그랑즈 함수가 정의되는 확장된 공간에서 안장점 (saddle point)이 되도록 다음 조건을 만족해야 한다.

$$\sum_j \beta_j = 1, \quad \alpha = \sum_j \beta_j \Phi(x_i), \quad \beta_j = C - \mu_j. \tag{2.1}$$

KKT (Karush-Kuhn-Tucker) 조건 (Kuhn과 Tucker, 1951)에 따라 다음과 같은 결과를 얻게 된다.

$$\xi_j \mu_j = 0, \quad (R^2 + \xi_j - \|\Phi(x_j) - \alpha\|^2)\beta_j = 0. \tag{2.2}$$

여기서 $\xi_j > 0$이고, $\beta_j > 0$인 점 x_j는 식 (2.2)에 따라 특징 공간 구의 바깥쪽에 놓여진다. $\xi_j \mu_j = 0$은 한 점이 $\mu_j = 0$을 가진다고 할 수 있으며, 식 (2.1)에 의해 $\beta_j = C$가 되는 자료를 경계서포트벡터 (bounded support vector; BSV)라고 한다. $\xi_j = 0$인 x_j는 특징공간의 구 표면이나 내부에 사상된다. 만약 한 점이 $0 < \beta_j < C$라면 특징공간의 구 표면에 놓여 있게 됨을 내포하며, 이를 서포트벡터 (support vector, SV)라고 한다. 서포트벡터들은 군집 경계 상에 놓이며, 경계서포트벡터들은 경계 밖에 놓인다. 그리고 다른 모든 점들은 경계 안에 존재하게 된다. $C \geq 1$일 때는 식 (2.1)에 의해 경계서포트벡터는 존재하지 않게 된다. 라그랑즈 함수를 변환하여 β_j로 표현하면 다음과 같다.

$$W = \sum_j \Phi(x_j)^2 \beta_j - \sum_{i,j} \beta_i \beta_j \Phi(x_i)\Phi(x_j).$$

여기서 $0 \leq \beta_j \leq C$, $j = 1, \cdots, n$이다. W는 Mercer 커널 (Mercer, 1909) 를 사용하여 다음과 같이 나타낼 수 있다.

$$W = \sum_j K(x_j, x_j)\beta_j - \sum_{i,j} \beta_i \beta_j K(x_i, x_j).$$

따라서 점 x에 대해 그 점이 특징공간에 사상된 상과 구 중심과의 거리를 구할 수 있다. 본 논문에서는 군집의 등고선을 산출하기 위해 가우시안 커널 (Gaussian kernel)을 사용하였으며, 식은 다음과 같다.

$$K(x_i, x_j) = e^{-q\|x_i - x_j\|^2}.$$

여기서 q는 너비 파라미터이다. X에 대해 그 점이 특징공간에 사상된 상과 구 중심과의 거리 R^2은 다음과 같다.

$$R^2(x) = \|\Phi(x) - \alpha\|^2.$$

식 (2.1)을 위의 식에 적용하면 다음과 같다.

$$R^2(x) = K(x_i, x_j) - 2\sum_j \beta_j K(x_j, x_j) + \sum_{i,j} \beta_i \beta_j K(x_i, x_j).$$

구의 반경 R은 서포트벡터로 구성되어 있으며 등고선들은 군집의 경계를 형성한다. 즉 서포트벡터는 군집 경계 위에 놓여 있으며, 경계서포트벡터는 경계 바깥쪽에, 나머지 다른 자료들은 군집 안쪽에 놓인다. 군집 서술자 (cluster description)는 다른 군집들에 속하는 점들을 구분 짓지는 못한다. 구분 짓기 위해서는 근접 행렬 $A_{i,j}$가 기하관측 (geometric observation)을 기반으로 정의되어야 한다. 다른 군집에 속해있는 자료 점들이 쌍으로 주어졌을 때, 자료 공간에서 어떤 패스가 연결되기 위해서는 특징 공간에서 대응되는 패스들이 구의 바깥쪽에서 교차해야만 한다. 쌍으로 주어진 자료 x_i, x_j에 대해 근접 행렬 $A_{i,j}$은 다음과 같은 이진 값을 가진다.

$$A_{i,j} = \begin{cases} 1, & \text{if } R(x_i + \lambda(x_j - x_i)) \leq R, \quad \forall \lambda \in [0,1]; \\ 0, & \text{otherwise.} \end{cases}$$

군집들은 A에 의해 안내된 그래프의 연결된 구성요소처럼 정의되었다. 자료 x_i, x_j에 대한 $A_{i,j}$을 계산할 때는 m번의 샘플링에 의해 수행되며, 일반적으로 샘플링 횟수는 10에서 20 사이이다. Ben-Hur 등 (2001a, 2001b)의 군집 수가 2인 군집 라벨링 전략 (cluster labeling strategy)은 다음과 같다.

군집 라벨링 전략 2 (cluster labeling strategy 2)는 전략 1보다 라벨링 속도가 더 빠르며, 전략 2에서는 n_{sv}는 $0.05n - 0.1n$보다 크다. 전략 1과 전략 2에서 소요되는 시간은 이차방정식을 풀어야 함에 따라 복잡하다. 자료 공간에서 가까운 군집 경계에 있는 자료 점들은 동일한 군집으로 묶이는 것을 더 선호한다. 군집 라벨링 결과는 3장의 실험 및 결과분석을 통해 확인할 수 있다.

Table 2.1 Cluster labeling strategy

cluster labeling strategy 1 (CG)
Calculate $A_{i,j}$ for each pair of points x_i and x_j in data space. This results in using the complete graph, denoted by CG(complete graph), to model adjacency $A_{i,j}$. It takes $O(n^2m)$ time.
cluster labeling strategy 2 (SVG)
Calculate $A_{i,j}$ only for pairs of points x_i and x_j, where x_i or x_j is a support vector. This results in a subgraph of CG, which is referred to as SVG. It takes $O((n - n_{bsv})n_{sv}^2 m)$ time, where n_{bsv} is the number of BSVs and n_{sv} is the number of free SVs.

3. 실험 및 결과분석

본 논문에서는 한국대학교육협의회 대학정보공시센터의 대학 알리미에서 공시되어 있는 대구·경북지역의 4년제 대학 24개교의 대학지표 자료를 가지고 군집분석을 통해 도출된 결과로 동일 군집별 대학들의 특징을 분석하고자 한다. 분석에서는 교육대와 산업대는 대학의 특수성에 따라 분석에서 제외시켰다. 공시된 대학지표 자료는 대학명, 소재, 구분, 분류, 연도, 최종등록률, 입학정원, 모집인원, 신입생경쟁률, 재학생충원률, 일반계고 비율, 전문계고 비율, 진학률, 건강보험 취업률, 전임1인당 학생 수 편제 정원, 전임 1인당 학생 수 재학생, 중도 탈락생 비율, 등록금, 등록금인상률, 학생1인당교육비, 재학생1인당장학금으로 구성되어 있다. 공시된 대학지표 자료 중에서 재학생충원률 (충원률)과 건강보험 취업률 (취업률)을 기준으로 대학들을 분류하고자 한다. 충원률과 취업률에 따른 대학별 군집분석을 위해 일반적으로 많이 사용하는 SPSS의 계층적 군집분석과 서포트백터 군집분석을 가지고 분류하였다.

먼저 군집분석을 위해 일반적으로 많이 사용하는 SPSS의 계층적 군집분석을 이용하여 계층적으로 군집들이 어떻게 분류되는지 덴드로그램을 작성한 후 결과를 분석한다. 군집을 형성하기 위해 군집 형성 개수를 2～4로 지정하였다. 결과는 Figure 3.1과 Table 3.1에 나타나 있다.

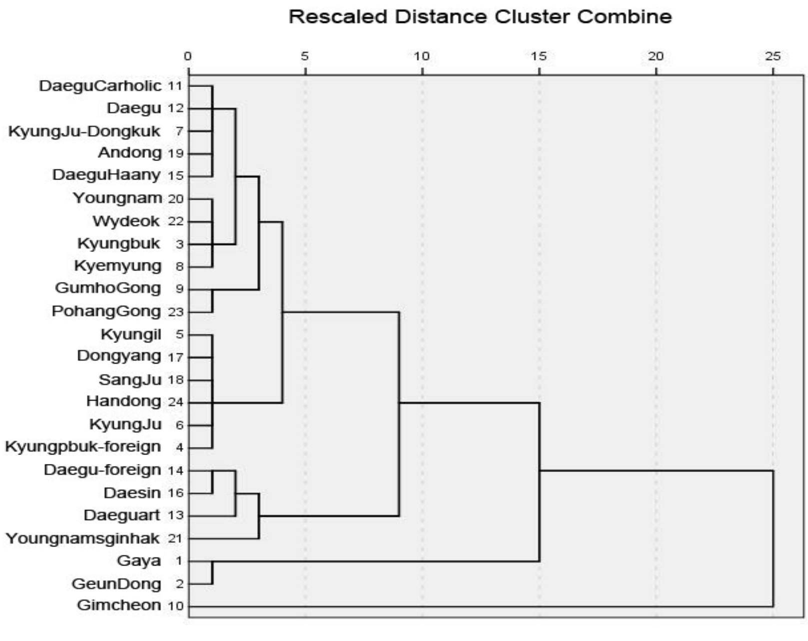

Figure 3.1 Hierarchical clustering analysis (dendrogram)

Figure 3.1은 대학별로 군집되어 지는 과정이 나와 있으며, Table 3.1에서는 군집의 개수가 4, 3, 2인 경우로 분리해서 정리되어 있다. 2 군집인 경우는 신생대학인 김천대만 다르게 분류되어 있다. 3 군집인 경우에는 가야대, 건동대가 군집 1로 분류되어 있으며, 김천대가 군집 3으로, 나머지 다른 대학들은 군집 2로 분류되어 있다. 4 군집인 경우에는 가야대, 건동대가 군집 1로 분류되어 있으며, 김천대가 군집 3으로, 대구외국어대, 영남신학대, 대구예술대가 군집 4로, 나머지 다른 대학들은 군집 2로 분류되어 있다. 군집 3으로 분류된 김천대는 신생대학이며, 대구외국어대와 영남신학대, 대구예술대는 외국어, 신학, 예술 분야로 특성화된 대학으로 일반 4년제 대학과는 다른 특성이 있음을 알 수 있다. 또한 가야대, 건동대는 4년제 대학이면서 전체 학과 수가 15개 이하로 소규모의 대학이며, 특정 분야만 특성화된 대학으로 분류됨을 알 수 있다.

Table 3.1 Result of hierarchical clustering analysis

case	cluster4	cluster3	cluster2
Gaya	1	1	1
GeunDong	1	1	1
Kyungbuk	2	2	1
Kyungpbuk-foreign	2	2	1
Kyungil	2	2	1
KyungJu	2	2	1
KyungJu-Dongkuk	2	2	1
Kyemyung	2	2	1
GumhoGong	2	2	1
Daegu	2	2	1
DaeguCarholic	2	2	1
Handong	2	2	1
SangJu	2	2	1
Andong	2	2	1
DaeguHaany	2	2	1
PohangGong	2	2	1
Dongyang	2	2	1
Youngnam	2	2	1
Wydeok	2	2	1
Gimcheon	3	3	2
Daesin	4	2	1
Daegu-foreign	4	2	1
Youngnamsginhak	4	2	1
Daeguart	4	2	1

다음으로 서포트백터 군집분석 기법을 이용하여 군집분석을 실시한다. 서포트백터 군집분석 기법으로는 Yang (2002)에 제안한 proximity graph 방식인 DD, MST, k-NN과 Lee와 Lee (2005)에서 제안한 dynamical system인 SEP-CG 방식을 사용하였다. 서포트백터 군집분석를 위해 사용된 커널 함수는 가우시안 커널 함수를 사용하였다. 다중 서포트백터 군집분석에서 모수인 벌칙상수 C와 커널 모수 σ의 최적 값 선택이 중요하며, 최적의 모수를 선택하기 위해 교차타당성 (cross validation; CV)를 이용하였다. 서포트백터 군집분석을 위해 사용된 커널 모수 σ와 벌칙상수 C값은 군집 수가 4 군집이 되도록 조절하였으며, 커널 모수 σ값과 벌칙상수 C값은 각각 0.9, 0.5로 구해졌다. 결과는 Figure 3.2의 좌측그림에서 Figure 3.3의 우측그림과 같으며, 각 대학별 소속 군집은 Table 4.1에 정리되어 있다. Figure 3.2에서 좌측그림은 DD 방식으로 군집분석한 결과이며, 우측그림은 MST 방식으로 군집분석한 결과이다. Figure 3.3에서 좌측그림은 k-NN 방식으로 군집분석한 결과이며, 우측그림은 SEP-CG 방식으로 군집분석한 결과이다. Figure 3.2에서 Figure 3.3를 보면 ○ 표기는 서포트벡터를 나타내며, × 표기는 이상치를 나타낸다. 각 대학별 충원률과 취업률에 따른 군집분석을 위해 기준이 되는 서포트벡터를 구한 다음에 그 서포트벡터를 기준으로 군집을 형성하게 된다.

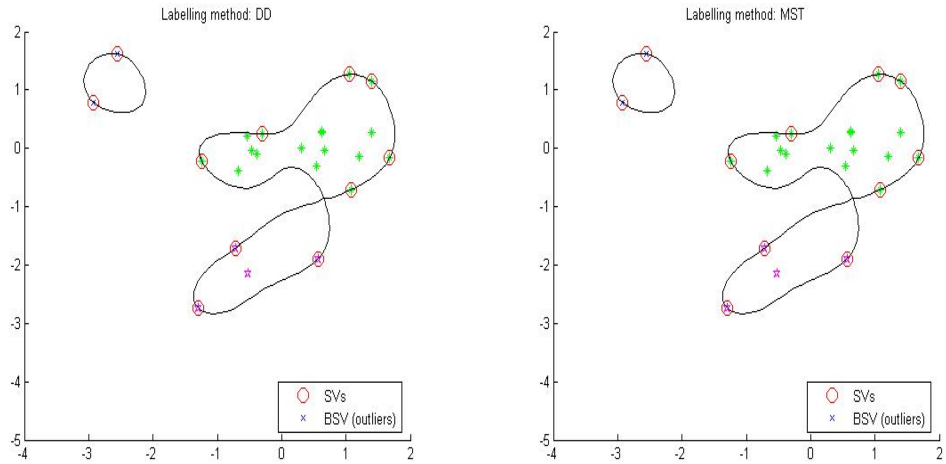

Figure 3.2 Clustering of DD method and clustering of MST method

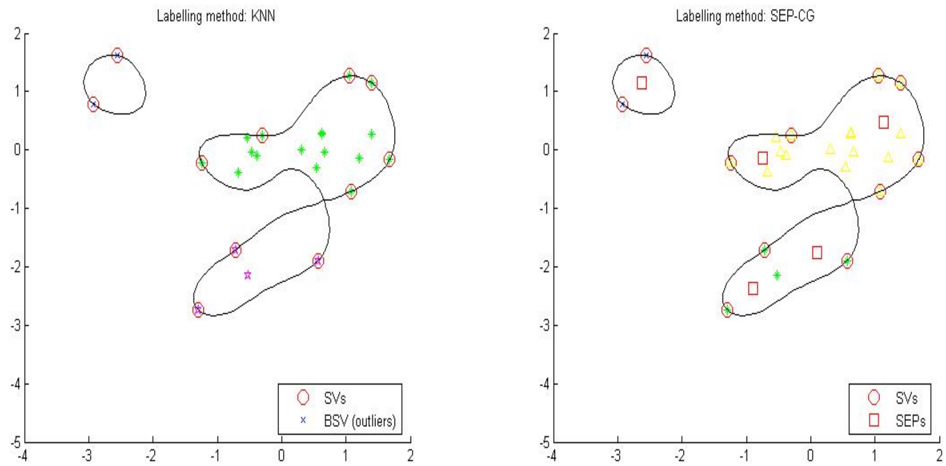

Figure 3.3 Clustering of k-NN method and clustering of SEP-CG method

4. 결론

본 논문에서는 그룹화된 대학들의 특징들을 살펴보기 위해 SPSS의 계층적 군집분석과 서포트백터 군집분석 기반의 방법들을 활용하여 분석한 후 공통으로 도출할 수 있는 정보를 분석하였다. 분류 결과는 Table 4.1에 정리되어 있다. Table 4.1를 보면 그룹 1에는 가야대와 건동대, 그룹 3에는 김천대, 그룹 4에는 대신대, 대구외대, 영남신학대, 대구예술대, 그룹 2에는 기타 나머지 대학으로 분류되어 있다. 그룹 1의 특징을 살펴보면 4년제 대학이면서 전체 학과 수가 15개 이하의 소규모 대학이며, 그룹 3에 해당하는 김천대는 신성대학이라는 특징이 있으며, 그룹 4의 대학들은 외국어, 신학, 예술 분야로 특성화

된 대학으로 일반 4년제 대학과는 다른 특성화 대학임을 알 수 있다. 그룹 2는 일반 4년제 대학 중 규모나 여러 가지 항목에서 조금은 안정적인 대학이라 할 수 있다. 각 그룹별로도 알 수 있듯이 학교마다 특성이 있어서 동일한 기준으로 대학을 판단하는 지표로 대학을 판단한다면 불이익이 있을 수 있는 요지가 있으므로 대학들을 평가한다든지 분류를 할 때는 그 특성을 최대한 반영할 수 있도록 고려할 필요성이 있다. 또한 이 분류 대학들 중에서 교과부에서 발표한 폐교로 선정된 대학도 있다. 사전에 판단하는 분류 기준이 명확하다면 그 분류 기준을 이용하여 군집분석을 이용한다면 대학마다 사전에 준비할 수 있는 사전 작업으로 활용할 수 있다.

Table 4.1 Clustering result

Univ.	SPSS	DD	MST	KNN	SEP-CG	SEP-CG(edit)
Gaya	1	1	1	1	1	1
Geondong	1	1	1	1	1	1
Kyungbuk	2	2	2	2	3	2
Kyungpbuk-foreign	2	2	2	2	3	2
Kyungil	2	2	2	2	3	2
KyungJu	2	2	2	2	3	2
KyungJu-Dongkuk	2	2	2	2	3	2
Kyemyung	2	2	2	2	3	2
GumhoGong	2	2	2	2	3	2
Daegu	2	2	2	2	3	2
DaeguCatholic	2	2	2	2	3	2
Handong	2	2	2	2	3	2
SangJu	2	2	2	2	3	2
Andong	2	2	2	2	3	2
DaeguHaany	2	2	2	2	3	2
PohangGong	2	2	2	2	3	2
Dongyang	2	2	2	2	3	2
Youngnam	2	2	2	2	3	2
Wydeok	2	2	2	2	3	2
Gimcheon	3	3	3	3	4	3
Daesin	4	4	4	4	2	4
Daegu-foreign	4	4	4	4	2	4
Youngnamsginhak	4	4	4	4	2	4
Daeguart	4	4	4	4	2	4

References

Ben-Hur, A., Horn, D., Siegelmann, H. T. and Vapnik, V. (2001a). Support vector clustring. *Journal of Machine Learning Research*, **2**, 125-137.

Ben-Hur, A., Horn, D., Siegelmann, H. T. and Vapnik, V. (2001b). A support vector method for hierarchical clustering. In *Advances in Neural Information Processing Systems*, **13**, MIT Press, New York.

Hwang, C. (2011). Asymmetric least squares regression estimation using weighted least squares support vector machine. *Journal of the Korean Data & Information Science Society*, **22**, 999-1005.

Hwang, C. and Shim, J. (2012). Mixed effects least squares support vector machine for survival data analysis. *Journal of the Korean Data & Information Science Society*, **23**, 739-748.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*, University of California Press, Berkeley, 481-492.

Lee, J. and Lee, D. (2005). An Improved cluster labeling method for support vector clustering. *IEEE Transactions on Pattern Aanalysis and Machine Intelligence*, **27**, 461-464.

Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society A*, 415-446.

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J. and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**, 1443-1472.

Shim, J. and Seok, K. (2013). GACV for partially linear support vector regression. *Journal of the Korean Data & Information Science Society*, **24**, 391-399.

Tax, D. M. J. and Duin, R. P. W. (1999). Support vector domain description. *Pattern Recognition Letters*, **20**, 1191-1199.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York.

Vapnik, V. (1998). *Statistical learning theory*, Springer, New York.

Yang, J., Estivill-Castro, V. and Chalup, S. K. (2002). Support vector clustering through proximity graph modelling. *Proceeding 9th International Conference, Neural Information Processing*, 898-903.

Journal of the Korean Data & Information Science Society
2013, **24**(4), 783–791

http://dx.doi.org/10.7465/jkdi.2013.24.4.783
한국데이터정보과학회지

Classification of universities in Daegu·Gyungpook by suppot vector cluster analysis

Hye Jung Park[1] · Jong Tae Kim[2]

[1]Common Liberal Arts, Daegu University
[2]Department of Computing & Statistics, Daegu University

Received 13 May 2013, revised 14 June 2013, accepted 1 July 2013

Abstract

There are sixteen indicators of "College Information" found on the website of College Information Disclosure Center. Among these indicators, the current study examined an enrollment rate and an employment rate based on health insurance coverage, and focused on twenty-four universities in Daegu and Gyeongbuk area. The universities were classified into groups by the enrollment rate and employment rate. This study investigated the characteristics pertaining to those different groups. Hierarchical cluster analysis and support vector cluster analysis were conducted in order to analyze the characteristics of the groups statistically.

Keywords: Cluster analysis, Gaussian kernel, support vector cluster.

[1] Invitation professor, Common Liberal Arts, Daegu University, Gyeongbuk 712-714, Korea.
[2] Corresponding author: Professor, Department of Computing & Statistics, Daegu University, Gyeongbuk 712-714, Korea. E-mail: jtkim@daegu.ac.kr