The Korean Journal of Applied Statistics (2020) DOI: http://dx.doi.org/10.5351/KJAS.2020.33.2.115
33(2), 115–122

Spectral clustering: summary and recent research issues

Sanghun Jeong[a] · Suhyeon Bae[a] · Choongrak Kim[a,1]

[a]Department of Statistics, Pusan National University

(Received February 19, 2020; Revised March 3, 2020; Accepted March 9, 2020)

Abstract

K-means clustering uses a spherical or elliptical metric to group data points; however, it does not work well for non-convex data such as the concentric circles. Spectral clustering, based on graph theory, is a generalized and robust technique to deal with non-standard type of data such as non-convex data. Results obtained by spectral clustering often outperform traditional clustering such as K-means. In this paper, we review spectral clustering and show important issues in spectral clustering such as determining the number of clusters K, estimation of scale parameter in the adjacency of two points, and the dimension reduction technique in clustering high-dimensional data.

Keywords: adjacency, dimension reduction, number of clusters, scale parameter

1. 서론

클러스터링(clustering)은 자료의 분석을 위한 여러 통계 기법 중 가장 널리 사용되는 것 중의 하나로서, 통계학은 물론 컴퓨터 과학, 생물학, 의학, 공학 등 다양한 분야에 많이 이용되고 있다 (von Luxburg, 2007). 특히, 인공지능을 위한 기계학습의 도구로서 인공 신경망을 이용한 분석에도 유용하게 사용되고 있다. 최근 인공지능 분야의 3대 석학인 LeCunn 등 (2015)은 Nature지에 게재한 Deep Learning이란 제목의 논문에서 딥 러닝의 미래는 클러스터링에 근거한 비지도 학습(unsupervised learning)이 지도 학습보다 훨씬 더 중요한 역할을 할 것으로 예측했다. 왜냐하면, 대부분의 인간과 동물의 사물에 대한 판단 및 구분 능력은 각 사물의 이름을 익혀서 인식하기보다 그 사물 자체를 보고 느낌으로써 인식하므로 지도 학습을 위한 분류보다 비지도 학습을 위한 클러스터링 문제가 훨씬 더 중요하기 때문이다. 예를 들어, 사람들이 있는 거리 풍경을 찍은 사진에서 사람과 배경을 구분짓는 것은 분류의 문제가 아니라 클러스터링의 문제이다. 또한, 자율주행차의 핵심기술인 영상인식에는 지도 학습보다 비지도 학습이 훨씬 더 많이 이용된다.

클러스터링의 목적은 주어진 자료를 몇 개의 그룹으로 나누는 것인데 같은 그룹 내의 자료들은 서로 유사하고 (유사도에 대한 metric은 2절에서 자세히 소개될 것임) 다른 그룹에 속하는 자료들은 상이하다는 조건을 만족해야 한다. 클러스터링은 매우 오랜 역사를 가진 통계적 분석법으로서 대부분의 다변량 통계학 분야에서 중요한 토픽으로 다루어지고 있다. 현재 많이 사용되고 있는 클러스터링 방법으로 K-평균 클러스터링(K-means clustering), 계층적 클러스터링(hierarchical clustering), 자기 조직화 지

This work was supported by a 2-year Research Grant of Pusan National University.
[1]Corresponding author: Department of Statistics, Pusan National University, Busandaehak-ro 63 beon-gil 2, Gumjeong-gu, Busan, 46288, Korea. E-mail: crkim@pusan.ac.kr

도(self-organizing map), 그리고 스펙트럴 클러스터링(spectral clustering)등이 있다. 이 중에서도 특히 K-평균 클러스터링이 가장 널리 사용되고 있으나 유사도(similarity)가 구면체(spherical) 또는 타원체(elliptical)로 정의되어 각 클러스터가 볼록 집합(convex set)의 형태인 자료에는 좋은 결과를 주지만 그렇지 않은 경우, 특히 자료가 다양체 구조(manifold structure)인 경우에는 매우 형편없는 결과를 나타낸다. 여러 형태의 클러스터링에 대한 체계적 논의는 대부분의 다변량 통계학 도서에 잘 정리되어 있는데 그 중에서도 Hastie 등 (2008)의 14장을 추천한다.

스펙트럴 클러스터링은 K-평균 클러스터링의 단점을 잘 보완해 줄 뿐 아니라 여러 형태의 자료나 고차원 자료(high-dimensional data) 등에 대해서도 좋은 결과를 나타내서 최근 인공 신경망 모형에는 대부분 스펙트럴 클러스터링이 이용되고 있다. 스펙트럴 클러스터링에 대한 평가 논문 중에서 von Luxburg (2007)이 많이 인용되고 있다. 본 논문에서는 최근 많은 연구자들의 관심을 받고 있는 스펙트럴 클러스터링에 대해 알기 쉽게 소개하고, 어떤 문제점이 있으며 최근의 연구 방향은 무엇인지 소개하고자 한다.

2. 스펙트럴 클러스터링

스펙트럴 클러스터링은 그래프 이론에 바탕을 두고 있으며, 알고리즘의 마지막 부분은 K-평균법을 이용한다. 먼저 군집화하고자 하는 n개의 자료를 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ 이라 하며 여기서 $\boldsymbol{x}_i$는 p-차원이다.

2.1. 그래프 이론

자료들간의 인접성(adjacency) 정도는 흔히 네트워크(network)로 표시할 수 있으며 이러한 네트워크는 그래프 $G = G(V, E)$로 나타낼 수 있다. 여기서, $V = \{1, \ldots, n\}$는 꼭짓점(vertex)의 집합이고, E는 $V \times V$상에 존재하는 변(edge)의 집합을 나타낸다. 흔히, i번째 꼭짓점은 p-차원 벡터인 $\boldsymbol{x}_i$, $i = 1, \ldots, n$로 표시한다. 이제 a_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n$를 i번째와 j번째 관측치 간의 유사도 또는 인접성을 나타낸다고 하자. 만약 $a_{ij} \neq a_{ji}$이면 방향성(directed) 그래프, $a_{ij} = a_{ji}$이면 비방향성(undirected) 그래프라 한다. 또한, a_{ij}가 0 또는 1의 값만 취하면 비가중 인접성(unweighted adjacency)이라 하고, 인접한 정도에 비례하여 여러 가지 값을 취하면 가중 인접성(weighted adjacency)이라 한다. 가중 인접성의 경우 주로 0과 1사의 값을 취하는 경우가 많은데 가장 널리 사용되는 가중 인접성은 가우시안 커널(Gaussian kernel)로서 $a_{ij} = \exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/c)$로 주어진다. 단, 여기서 c는 척도(scale)를 나타내는 모수로서 자료로 부터 추정하여 사용한다. 그 외에도 두 관측치의 표본 상관계수의 절댓값이나 제곱한 것을 사용하기도 한다. 본 논문에서는 비방향성이고 가중 인접성을 가정한다. 방향성이 있거나, 가중인접성을 가정하지 않는 자료는 매우 드물기 때문이다. 한편, i번째 관측치의 인접성 정도(degree)는 $d_i = \sum_{j=1}^{n} a_{ij}$로 주어지는데 이는 i번째 관측치와 다른 관측치들 간의 인접성을 모두 합한 것이다. 아울러, 각 관측치의 인접성 정도를 대각원소로 갖는 행렬 $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_n)$를 인접성 정도행렬(degree matrix)이라 하며, 인접성 정도행렬에서 인접 행렬을 뺀 행렬 $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$를 라플라스 행렬(Laplacian matrix)이라 부른다. 라플라스 행렬은 스펙트럴 클러스터링에서 매우 중요한 역할을 하는데 이와 관련된 성질은 다음과 같다. 만약 주어진 네트워크가 K개의 완전분리된 그룹(fully separated group)일 경우, 즉 그룹간 모든 인접 $a_{ij} = 0$인 경우, 라플라스 행렬의 차수(rank)는 $n - K$가 된다는 사실이다. 예를 들어 가우시안 커널이 인접성 척도로 사용되면 주어진 네트워크는 모든 관측치들간의 인접성이 양수이기 때문에 하나의 그룹으로 나타나므로 이 경우 라플라스 행렬의 차수는 $n - 1$이 된다. 또한, 라플라스 행렬은 양반정치(positive semi-definite)이므로 모든 고유치(eigenvalue)가 0보다 크거나 같다. 따라서, 가우시안 커널을 인접성 척도로 사용한 네트워크의 경우 0인 고유치는 한 개, 나머지 $n - 1$개의 고유치는 모두 양수이다. 이 중에서도 0이 아닌 가장 작은 양수의 고유치에 해당되는 고유벡터(eigenvector)를 특별히 피들러 벡터(Fiedler vector)라 하고, 이는 클러스터링과 밀접한

관계가 있는 것으로 알려져 있다 (Fiedler, 1973; Kim 등, 2008).

2.2. 스펙트럴 클러스터링

스펙트럴 클러스터링과의 비교를 위해 먼저 K-평균 클러스터링을 간략하게 소개한다. K-평균 클러스터링은 클러스터링 중 가장 많이 사용되는 것으로 관측치를 구성하는 p개의 변수 모두 양적 변수(quantitative)이며 각 클러스터의 형태가 볼록 집합일 경우 매우 만족스러운 결과를 나타낸다. 왜냐하면, K-평균 클러스터링에 사용되는 거리측도는 유클리드 거리(Euclidean distance)의 제곱인 $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2$으로 주어지기 때문이다. 앞에서도 언급했듯이 각 클러스터의 형태가 볼록 집합이 아닌 경우에는 매우 만족스럽지 못한 결과를 주며 (Zelnik-Manor과 Perona (2005)의 그림 1에 여러 형태의 예제가 잘 나타나 있음), 특히 K가 변하면 클러스터링 결과 자체가 완전히 다른 모습으로 나타나서 클러스터의 갯수에 따른 형태적 연속성이 전혀 없다. 더불어, 초기 평균값의 설정에도 매우 민감하다.

스펙트럴 클러스터링의 아이디어는 Fiedler (1973)에 의해 처음으로 제안되었는데 그래프를 이용한 네트워크에서 스펙트럴 분석에 의해 인접행렬의 고유치와 고유벡터가 클러스터링에 이용되기 때문에 이런 이름이 붙여졌다. 구체적으로 보면 주어진 K에 대하여 클러스터링에 대한 정보를 가진 K개의 직교정규화 벡터 (orthonormal vectors) $\mathbf{f}_1, \ldots, \mathbf{f}_K$로 구성된 미지의 $n \times K$ 행렬 $\boldsymbol{F}$가 있을 때, 스펙트럴 클러스터링은

$$\min_{\boldsymbol{F}} \operatorname{tr}\left(\boldsymbol{F}^T \boldsymbol{L} \boldsymbol{F}\right) \quad \text{s.t.} \quad \boldsymbol{F}^T \boldsymbol{F} = \boldsymbol{I}$$

에 의해 $\boldsymbol{F}$를 추정하는 것이다. Ng 등 (2002)에 의해 제안된 정규화 스펙트럴 클러스터링(normalized spectral clustering)이 사용되기 이전에는 비정규화 스펙트럴 클러스터링(unnormalized spectral clustering)이 주로 사용되고 있었는데 두 가지 방법의 알고리즘은 각각 다음과 같다.

Algorithm 1 비정규화 스펙트럴 클러스터링

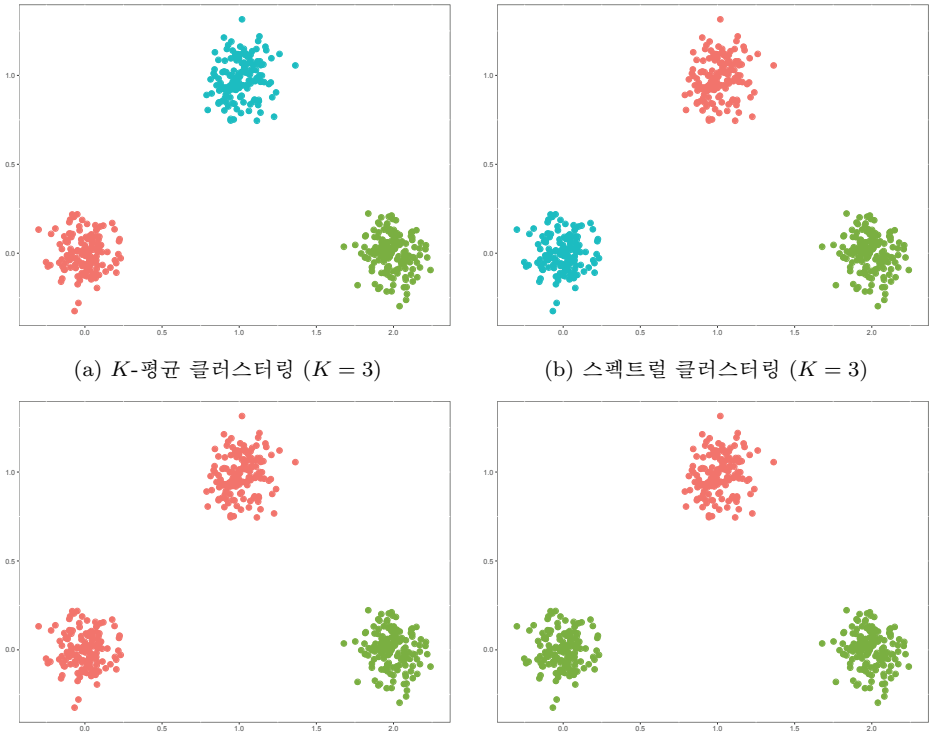
단계1. $\boldsymbol{y}_i, i = 1, \ldots, n$를 계산한다. 단, $\boldsymbol{y}_i$는 $n \times K$ 행렬 $U = (\mathbf{u}_1, \ldots, \mathbf{u}_K)$의 i번째 행이고, $(\mathbf{u}_1, \ldots, \mathbf{u}_K)$는 라플라스 행렬 $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$의 n개의 고유벡터 중 가장 작은 K개의 고유벡터들이다.
단계2. n개의 벡터 $\mathbf{y}_i, \ldots, \mathbf{y}_n$에 K-평균 클러스터링을 적용한다.

Algorithm 2 정규화 스펙트럴 클러스터링

단계1. $\boldsymbol{y}_i, i = 1, \ldots, n$를 계산한다. 단, $\boldsymbol{y}_i$는 $n \times K$ 행렬 $U = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$의 i번째 행이고, $(\mathbf{u}_1, \ldots, \mathbf{u}_K)$는 정규화 라플라스 행렬 $\boldsymbol{L}^* = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{L} \boldsymbol{D}^{-\frac{1}{2}}$의 n개의 고유벡터 중 가장 작은 K개의 고유벡터들이다.
단계2. $n \times k$ 행렬 $\boldsymbol{U}$의 각 행의 크기 (norm)가 1이 되도록 행렬 $\boldsymbol{T}$를 만든다.
단계3. 행렬 $\boldsymbol{T}$의 n개의 행 $(\mathbf{t}_1, \ldots, \mathbf{t}_n)$에 K-평균 클러스터링을 적용한다.

3. 예제 및 향후 연구과제

K-평균 클러스터링과 스펙트럴 클러스터링의 단점을 지적하는 예를 들기 위해 두 가지 형태의 가상적인 자료를 이용한다. 군집분석에서 클러스터링의 효율성은 흔히 오분류율(misclassification rate)을 기준으로 한다.



(a) K-평균 클러스터링 ($K = 3$) (b) 스펙트럴 클러스터링 ($K = 3$)

(c) K-평균 클러스터링 ($K = 2$) (d) 스펙트럴 클러스터링 ($K = 2$)

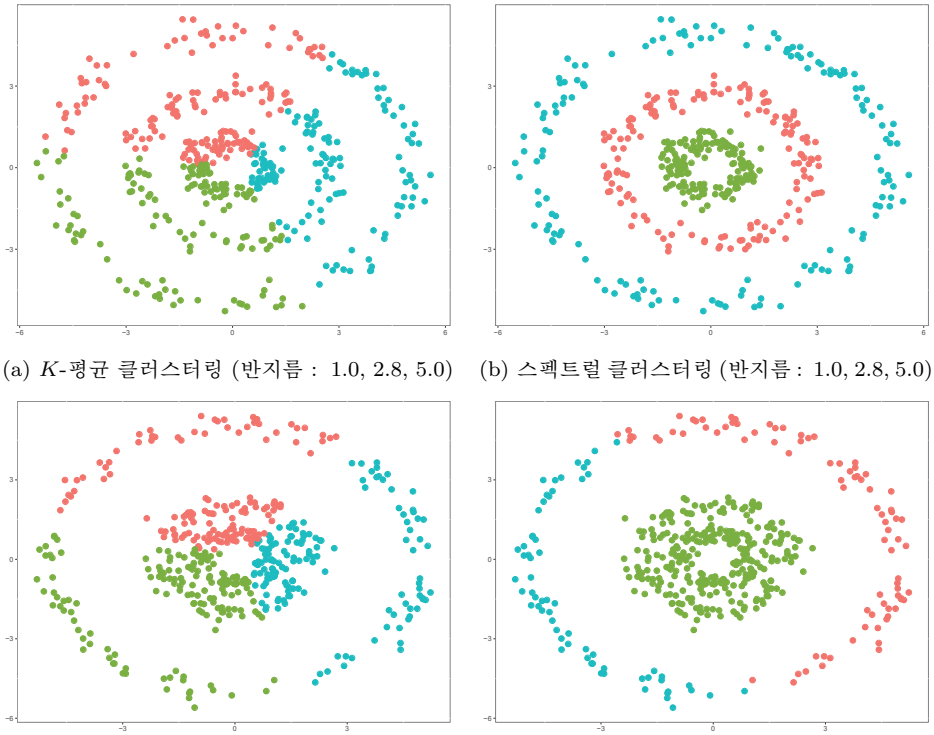
Figure 3.1. Data with type of convex set; K-means and spectral clusterings are applied when $K = 2$ and 3, respectively.

3.1. 볼록 집합 형태의 자료

Figure 3.1의 자료는 볼록 집합 형태의 자료로서 K-평균 클러스터링과 스펙트럴 클러스터링이 제대로 클러스터링 할 수 있는 것이지만 클러스터의 갯수 K가 잘못 지정되거나 추정되었을 경우의 심각성을 나타내는 자료이다. 이를 위해 $\mathrm{E}(X_1) = (\mu_1, \mu_2)^T, \mathrm{Cov}(X_1) = \sigma^2 I_2$인 이변량 정규분포로 부터 $\sigma = 0.1$, 0.3에 대해 $X_1, \ldots, X_{150}$은 $\mu_1 = \mu_2 = 0$일때, $X_{151}, \ldots, X_{300}$은 $\mu_1 = 2, \mu_2 = 0$일때, 그리고 $X_{301}, \ldots, X_{450}$은 $\mu_1 = \mu_2 = 1$일때 임의표본을 생성하였다. Figure 3.1의 (a), (b)에서 보이는 것 처럼 $K = 3$으로 제대로 지정된 경우 K-평균 클러스터링과 스펙트럴 클러스터링 모두 제대로 된 결과를 나타내지만, Figure 3.1의 (c), (d)와 같이 $K = 2$로 잘못 지정된 경우 두 클러스터링 모두 잘못된 결과를 나타낸다.

3.2. 동심원 자료

Figure 3.2의 자료는 동심원 자료로서 3개의 동심원에 각각 150개의 점들이 있는데 각 점들은 구간 $[0, 2\pi]$에서 연속균등분포로 부터 생성하였다. 반지름은 각각 1.0, 2.8, 5.0을 갖는 그룹 (Figure 3.2 (a), (b))과 1.0, 2.0, 5.0을 갖는 그룹 (Figure 3.2 (c), (d))이며 각 점들은 평균이 0, 표준편차가 0.25인 정규분포로 부터 추출된 오차를 갖는다. Figure 3.2 (a), (b)에서는 K-평균 클러스터링이 클러스터의 갯수 K가 3으로 주어져 제대로 지정되었음에도 불구하고 볼록 집합의 형태가 아닌 경우 만족스럽지 못



(a) K-평균 클러스터링 (반지름 : 1.0, 2.8, 5.0) (b) 스펙트럴 클러스터링 (반지름 : 1.0, 2.8, 5.0)

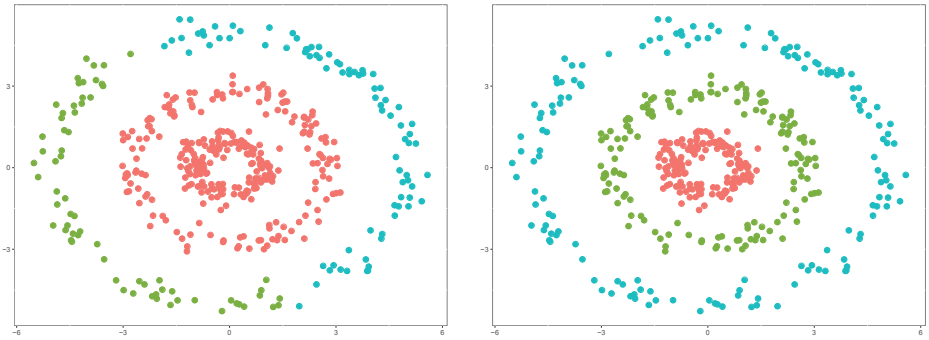
(c) K-평균 클러스터링 (반지름 : 1.0, 2.0, 5.0) (d) 스펙트럴 클러스터링 (반지름 : 1.0, 2.0, 5.0)

Figure 3.2. Data with type of concentric circles; K-means and spectral clusterings are applied when radii are 1.0, 2.8, 5.0, and 1.0, 2.0, 5.0, respectively.

한 결과를 나타내지만 스펙트럴 클러스터링은 좋은 결과를 나타내는 현상을 쉽게 확인할 수 있다. 한편, Figure 3.2 (c), (d)에서는 스펙트럴 클러스터링의 인접행렬에 사용된 커널의 척도 모수가 잘못 지정되거나 추정되었을 경우의 심각성을 볼 수 있다. 즉, Figure 3.2 (d)는 같은 형태의 자료이지만 반지름이 달라진 경우에는 결과가 매우 좋지록 못하며 척도모수가 제대로 추정되어야 함을 여실히 보여주고 있다.

3.3. 향후 연구과제

첫째, 클러스터의 갯수에 대한 사전 정보가 주어지는 경우는 예외지만 클러스터링에서 가장 어려운 미해결 문제(open problem)는 클러스터의 갯수 K의 추정이다. 스펙트럴 클러스터링은 여러 가지 상황에서 가장 안전하게 사용할 수 있는 방법이지만 다른 클러스터링과 마찬가지로 K의 추정은 피해갈 수 없다. K의 추정에 관한 연구 논문으로는 크게 모수적인 방법과 비모수적인 방법으로 나눌 수 있다. 모수적인 방법의 대표적 연구로는 Fraley와 Raftery (2002)가 제안한 로그-우도함수를 이용한 것으로 이는 자료들이 특정 분포로 부터 추출되었다는 가정하에 이루어지는 것이다. 비모수적 방법으로는 Tibshirani 등 (2001)가 scree plot을 이용하여 갭 통계량(gap statistic)을 제안하였고, Le와 Levina (2015)는 베테 헤시안 행렬(Bethe Hessian matrix)을 이용하여 클러스터의 갯수에 대한 추정을 제안하였다. 하지만, 이 모든 방법들은 매우 제한된 조건 하에 사용 가능하며 제약조건을 벗어난 경우에는 효율성이 떨어지는 단점이 있다. 예를 들어, 갭 통계량은 K-평균 클러스터링에는 비교적 좋은 결과를 보여주지만 스펙트럴



(a) 스펙트럴 클러스터링 (척도모수 : 1.0) (b) 스펙트럴 클러스터링 (척도모수 : 3.0)

Figure 3.3. Data with type of concentric circles; spectral clusterings are applied when scale parameters of the Gaussian kernel function are 1.0 and 3.0, respectively.

클러스터링에는 적용하기 어려운 것으로 알려져 있다.

둘째, 스펙트럴 클러스터링에서 해결되어야 할 또 다른 문제는 커널함수에서 척도모수에 대한 추정이다. 동심원 자료의 예에서 본 것 처럼 척도모수의 추정이 잘못되면 클러스터링의 결과 또한 매우 나쁘게 나타난다. Figure 3.3은 동심원 자료에서 가우시안 커널의 척도모수에 대한 추정이 얼마나 중요한지 보여주는 예다. 이를 위해 Zelnik-Manor과 Perona (2005)가 국소 추정법을 제안하였으나 적용할 수 있는 경우가 매우 제한적이고 일반적으로 사용할 수 있는 추정치가 될 수 없어서 추가 연구가 반드시 되어야 할 분야이다.

마지막으로, 지금 현재 스펙트럴 클러스터링의 연구에서 가장 활발하게 진행되는 분야는 바로 고차원 자료의 차원축소에 관한 것이다. 차원축소는 반드시 원래 자료의 다양체 구조를 반영할 수 있어야 하는데 이를 위해 많은 연구들이 진행중이다. 차원축소를 위한 방법으로 주성분분석(principal component analysis)을 가장 많이 이용하고 있으며, 대표적인 연구들로 Ben-Hur 등 (2001), Xu 등 (2005), Zhang 등 (2009), Nie 등 (2011), Zhou 등 (2013), Wang 등 (2019) 등이 있다.

4. 결론

스펙트럴 클러스터링은 현재 사용되고 있는 여러 클러스터링 방법중에 가장 널리 사용되고 있으며 특히 인공 신경망 분야에서 활용도가 매우 높다. 특히, 자료의 형태가 볼록 집합이 아닌 경우 K-평균 클러스터링은 매우 좋지 못한 결과를 나타내며 이에 대한 대안으로 스펙트럴 클러스터링이 많이 이용되고 있다. 하지만, 스펙트럴 클러스터링 또한 시급히 개선되어야 할 부분이 많다. 클러스터의 갯수에 대한 추정, 인접행렬의 구성시에 필요한 척도모수의 추정, 고차원 자료에서의 차원축소 방법 등이 해결되어야 한다. 본 논문에서는 스펙트럴 클러스터링에 대해 간략하게 소개하고 향후 해결되어야할 연구분야 등을 살펴보았다.

References

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

Fiedler, M. (1973). Algebraic connectivity of graphs, *Czechoslovak Mathematical Journal*, **23**, 298–305

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation, *Journal of the American Statistical Association*, **97**, 611–631.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning* (2nd Ed), Springer, New York.

Kim, C., Cheon, M., Kang, M., and Chang, I. (2008). A simple and exact Laplacian clustering of complex networking phenomena: Application to gene expression profiles. In *Proceedings of the National Academy of Science*, **105**, 4083–4087.

Le, C. M. and Levina, E. (2015). Estimating the number of components in networks by spectral methods, arXiv 1507.00827.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning, *Nature*, **521**, 436–444.

Ng, A., Jordan, M., and Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. *Advances in Neural Information Processing Systems, 849?856*, MIT Press.

Nie, F., Zeng, Z., Tsang, I. W., Xu, D., and Zhang, C. (2011). Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering, *IEEE Transactions on Neural Networks*, **22**, 1796–1808.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a dataset via the gap statistic. *Journal of the Royal Statistical Society* Series B, **63**, 411–423.

von Luxburg, U. (2007). A tutorial on spectral clustering, *Statistics and Computing*, **17**, 395–416.

Wang, Q., Qin, Z., Nie, F., and Li, X. (2019). Spectral embedded adaptive neighbors clustering, *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 1265–1271.

Xu, L., Neufeld, J., Larson, B., and Schuurmans, D. (2005). Maximum margin clustering, *Advances in Neural Information Processing Systems*, 1537–1544.

Zhang, K., Tsang, I. W., and Kwok, J. T. (2009). Maximum margin clustering made practical, *IEEE Transactions on Neural Networks*, **20**, 583?596.

Zelnik-Manor, L. and Perona, P. (2005). Self-tuning spectral clustering, *Advances in Neural Information Processing Systems, 1601–1608*, MIT Press.

Zhou, G. T., Lan, T., Vahdat, A., and Mori, G. (2013). Latent maximum margin clustering, *Advances in Neural Information Processing Systems*, 28–36.

스펙트럴 클러스터링 - 요약 및 최근 연구동향

정상훈[a] · 배수현[a] ·김충락[a,1]

[a]부산대학교 통계학과

(2020년 2월 19일 접수, 2020년 3월 3일 수정, 2020년 3월 9일 채택)

요 약

K-평균 클러스터링은 매우 널리 사용되고 있으나 유사도가 구면체 또는 타원체로 정의되어 각 클러스터가 볼록 집합 형태인 자료에는 좋은 결과를 주지만 그렇지 않은 경우에는 매우 형편 없는 결과를 나타낸다. 스펙트럴 클러스터링은 K-평균 클러스터링의 단점을 잘 보완해 줄 뿐아니라 여러 형태의 자료나 고차원 자료 등에 대해서도 좋은 결과를 나타내서 최근 인공 신경망 모형에 많이 이용되고 있다. 하지만, 개선되어야 할 단점도 여전히 많다. 본 논문에서는 스펙트럴 클러스터링에 대해 알기 쉽게 소개하고, 클러스터 갯수의 추정, 척도모수의 추정, 고차원 자료의 차원축소 등 스펙트럴 클러스터링에 대한 최근의 연구 동향을 소개한다.

주요용어: 인접성, 차원축소, 클러스터의 갯수, 척도모수.

본 연구는 부산대학교 2년 과제 연구비에 의하여 수행되었음.

[1]교신저자: (46288) 부산시 금정구 부산대학로 63번길 2, 부산대학교 통계학과. E-mail: crkim@pusan.ac.kr