비선형비부값스펙트르무리짓기에서 핵함수의 한가지 구성방법

김진성, 강성혁

우리는 자료처리에서 많이 리용되고있는 비부값스펙트르무리짓기에 대하여 연구하였다.

선행연구[1－3]에서는 비부값행렬분해의 목적함수에 정규화항을 추가하는 방법으로 무리짓기의 성능을 개선하였으며 선행연구[4]에서는 핵에 기초한 방법으로 선형비부값스펙트르무리짓기를 비선형인 경우로 확장하였다. 그러나 이 방법들에서는 류사도행렬을 표본벡토르들사이의 내적에 의해 정의하거나 또는 핵에 의해 일반화하여 정의하는 경우에도 어떤 핵함수를 리용해야 하는가에 대해서는 밝히지 못하였다.

론문에서는 여러개의 핵함수를 리용하여 주어진 자료에 적합한 새로운 핵함수를 구성하기 위한 방법과 그것을 리용한 비선형비부값스펙트르무리짓기알고리듬을 제기한다.

$X=(x_1,\ x_2,\ \cdots,\ x_n)\in\mathbf{R}^{m\times n}$ 을 표본자료행렬, $x_i\in\mathbf{R}^m$ 을 표본벡토르, W 를 류사도행렬이라고 하자. 즉 W_{ij} 는 x_i 와 x_j 사이의 류사도이다. 일반성을 잃지 않고 x_i 의 모든 성분들이 0 이상이며 표본들을 k 개의 무리로 가른다고 하자.

D 를 $D_{ii}=\sum_{j=1}^{n}W_{ij}$ 인 대각선행렬로 정의한다.

$\Phi: x_i\mapsto\Phi(x_i)$ 는 표본벡토르를 고차원특징공간으로 넘기는 넘기기로서 $\Phi(x_i)\geq 0$ 이라고 하고 Φ 에 대응한 핵함수를 K 라고 하면 고차원공간에서의 표본벡토르들사이의 류사도행렬은 $W_{ij}=K(x_i,\ x_j)=\langle\Phi(x_i),\ \Phi(x_j)\rangle=\Phi^{\mathrm{T}}(x_i)\Phi(x_j)$, $W=\Phi^{\mathrm{T}}(X)\Phi(X)$ 로 정의된다.

$H=(h_1/\|h_1\|,\ h_2/\|h_2\|,\cdots,\ h_k/\|h_k\|)$ 를 지시자행렬, h_l 을 무리 C_l $(l=1,\cdots,\ k)$ 에 대한 n 차원지시자벡토르라고 하자.

비선형비부값스펙트르무리짓기를 위해서는 다음의 최량화문제를 풀어야 한다.[1, 4]

$$\min_{H\geq 0,\ H^{\mathrm{T}}H=I}\|W-HH^{\mathrm{T}}\|_F^2,\quad \min_{H\geq 0,\ Z^{\mathrm{T}}Z=I}\|D^{-1/2}WD^{-1/2}-ZZ^{\mathrm{T}}\|_F^2 \tag{1}$$

여기서 H 는 RCut를 리용할 때의, Z 는 NCut를 리용할 때의 무리지시자행렬이다.

류사도행렬 W 는 핵함수 K 의 선택에 따라 달라진다. 또한 최량화문제 (1)의 목적함수값은 W 가 얼마나 잘 분해될수 있는가 하는것을 나타낸다고 할수 있다.

특수한 경우 목적함수값이 0 으로 되면 무리지시자행렬로부터 본래의 류사도행렬을 완전히 재구성할수 있으며 이것은 이때의 W 에 무리짓기에 필요한 자료의 특성이 모두 반영된것으로 생각할수 있다. 그러므로 최량화문제 (1)의 목적함수값이 작아질수록 핵함수 K 가 무리짓기에 보다 효과적이라고 할수 있다.

이로부터 론문에서는 여러개의 핵함수들을 결합하여 무리짓기에 보다 효과적인 핵함수를 구성하기 위한 방법을 제기한다.

$K_1,\ K_2,\ \cdots,\ K_s$ 를 핵함수들, 그것에 대응한 류사도행렬을 $W_1,\ W_2,\ \cdots,\ W_s$ 라고 하자.

이때 새로운 핵함수 K 를 다음과 같이 구성한다.

$$K=\alpha_1K_1+\alpha_2K_2+\cdots+\alpha_sK_s,\ \ \alpha_1+\alpha_2+\cdots+\alpha_s=1,\ \ \alpha_i\geq 0,\ \ i=1,\ 2,\cdots,\ s$$

K에 대응한 류사도행렬은 $W=\alpha_1 W_1+\alpha_2 W_2+\cdots+\alpha_s W_s$로 된다.

이처럼 류사도행렬을 결합하기 위해서는 먼저 W_i들을 표준화하여야 한다.

표시를 간단히 하기 위하여 하나의 핵함수에 의하여 얻어진 류사도행렬 W_i를 W로 표시하고 이것을 표준화하자.

쉽게는 $W'_{ij}=W_{ij}/\sqrt{p_i p_j}$ 형태의 표준화방법을 생각할수 있다. 여기서 $p_1,\ p_2,\ \cdots,\ p_n>0$은 표준화를 위한 상수들이다.

이때 새로 얻어진 류사도행렬 W'도 핵함수에 의해 얻어진 류사도행렬로 된다.

$p_1,\ p_2,\ \cdots,\ p_n>0$들을 잘 선택하면 류사도행렬이 특이한 성질을 만족하도록 표준화할수 있다.

방법 1 $p_i=W_{ii},\ 1\le i\le n$으로 선택하면 대각선원소들이 1이 되게 할수 있다.

또한 류사도행렬의 모든 원소들의 합이 1이 되도록 하는것은 아주 쉬우며 더우기는 류사도행렬에서 행의 원소들의 합과 렬의 원소들의 합이 모두 1이 되게 할수도 있다.

이것은 다음의 정리로부터 알수 있다.

정리 W를 대각선원소가 정수인 비부값대칭반정값행렬이라고 하자.

그러면 $p_1,\ p_2,\ \cdots,\ p_n>0$이 존재하여 $W'_{ij}=W_{ij}/\sqrt{p_i p_j}$로 정의되는 새로운 행렬 W'는 행의 원소들의 합이 1인 대칭반정값행렬로 된다.

증명 $W^0:=W$이고 $k\ge 0$에 대하여 W^k가 대각선원소가 정수인 비부값대칭반정값행렬이면 $d_i^k:=\sum_{j=1}^{n}W_{ij}^k$이고 D^k는 $D_{ii}^k=d_i^k$인 대각선행렬이며 $W^{k+1}:=(D^k)^{-1/2}W_k(D^k)^{-1/2}$이라고 하면 W^{k+1}도 대각선원소가 정수인 비부값대칭반정값행렬이다.

$W^k=(E^k)^{-1/2}W(E^k)^{-1/2}$으로 표시된다는것은 분명하다. 여기서 $E^k=D^{k-1}D^{k-2}\cdots D^0$.

그러면 W^k는 대칭반정값행렬이며 k가 증가함에 따라 행의 원소들의 합은 1로 수렴하게 된다. 그리고 대각선행렬 P와 첨수모임 T가 있어서 $\lim_{k\to\infty,\ k\in T}E^k=P$가 성립되며 또한 P의 대각선원소들을 $p_1,\ p_2,\cdots,\ p_n$이라고 하면 정리의 결론이 성립된다.(증명끝)

방법 2 정리의 증명과정에서 제시한 방법대로 류사도행렬 W를 행의 원소들의 합이 1에 충분히 가깝도록 표준화할수 있다.

류사도행렬 $W_1,\ W_2,\ \cdots,\ W_s$들을 우의 방법들로 표준화한다.

RCut를 리용한 스펙트르무리짓기에서는 류사도행렬의 균형을 맞추기 위하여 방법 1로, NCut를 리용한 스펙트르무리짓기에서는 방법 2로 표준화하는데 정리의 증명과정에서 알수 있는바와 같이 이렇게 하면 표준화후에도 그전의 목적함수와 류사하다고 할수 있고 계산에도 아주 편리하다.

류사도행렬들을 표준화하고 최량화문제 (1)에 대입하면 다음과 같다.

$$\min_{\alpha_1,\alpha_2,\cdots,\alpha_s}\ \min_{H}\|(\alpha_1W_1+\alpha_2W_2+\cdots+\alpha_sW_s)-HH^{\mathrm{T}}\|_F^2,\ \ H^{\mathrm{T}}H=I,\ \alpha_1+\alpha_2+\cdots+\alpha_s=1,\ \ H,\ \alpha\ge 0$$

$$\min_{\alpha_1,\alpha_2,\cdots,\alpha_s}\ \min_{Z}\|(\alpha_1W_1+\alpha_2W_2+\cdots+\alpha_sW_s)-ZZ^{\mathrm{T}}\|_F^2,\ \ Z^{\mathrm{T}}Z=I,\ \ \alpha_1+\alpha_2+\cdots+\alpha_s=1,\ \ Z,\ \alpha\ge 0$$

여기서 $\alpha=(\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_s)$이다.

이와 같이 류사도행렬들을 표준화하면 RCut와 NCut의 최량화문제는 동일하게 된다.

$$\min_{\alpha_1,\alpha_2,\cdots,\alpha_s}\ \min_{Y,\ F}\|(\alpha_1W_1+\alpha_2W_2+\cdots+\alpha_sW_s)-YF\|_F^2,\quad FY=I,\ Y^{\mathrm{T}}=F,\ \alpha_1+\cdots+\alpha_s=1,\ F,\ Y,\ \alpha\geq 0$$

이 문제를 푸는 대신 정규화파라메터 $\mu,\ \gamma>0$을 리용하여 다음의 문제를 풀자.

$$\min_{\alpha_1,\alpha_2,\cdots,\alpha_s,\ Y,\ F\geq 0}\|(\alpha_1W_1+\alpha_2W_2+\cdots+\alpha_sW_s)-YF\|_F^2/2+$$
$$+\mu(\alpha_1+\alpha_2+\cdots+\alpha_s-1)^2/2+\gamma(\|FY-I\|_F^2+\|Y^T-F\|_F^2)/2$$
$$L(\alpha,\ Y,\ F):=\|(\alpha_1W_1+\alpha_2W_2+\cdots+\alpha_sW_s)-YF\|_F^2/2+$$
$$+\mu(\alpha_1+\alpha_2+\cdots+\alpha_s-1)^2/2+\gamma(\|FY-I\|_F^2+\|Y^T-F\|_F^2)/2$$

우리는 선행연구[4]에서와 류사하게 국부최소점을 구하기 위하여 다른 변수들은 고정하고 α와 Y, F에 대하여 차례로 갱신하는 방법을 리용한다.

그라디엔트하강법을 적용하면 다음과 같이 갱신된다.

$$\alpha_i=\alpha_i-\eta_i\{[\mathrm{tr}(K_iW)+\mu(\alpha_1+\alpha_2+\cdots+\alpha_s)]-[\mathrm{tr}(K_iYF)+\mu]\}\ \ (\eta_i>0\text{은 상수})$$

매 갱신에서 $\alpha_i\geq 0$이 만족되도록 갱신하기 위하여 선행연구[4]에서와 류사하게 $\eta_i=\alpha_i/[\mathrm{tr}(K_iW)+\mu(\alpha_1+\alpha_2+\cdots+\alpha_s)]$로 설정하면 다음과 같은 갱신규칙을 얻는다.

$$\alpha_i=\alpha_i[\mathrm{tr}(K_iYF)+\mu]/[\mathrm{tr}(K_iW)+\mu(\alpha_1+\alpha_2+\cdots+\alpha_s)] \tag{2}$$

마찬가지로 Y와 F에 대한 갱신규칙을 다음과 같이 얻는다.

$$Y_{ij}=Y_{ij}\frac{[WF^{\mathrm{T}}+2\gamma F^{\mathrm{T}}]_{ij}}{[YFF^{\mathrm{T}}+\gamma(F^{\mathrm{T}}FY)+\gamma Y]_{ij}},\quad F_{ij}=F_{ij}\frac{[Y^{\mathrm{T}}W+2\gamma Y^{\mathrm{T}}]_{ij}}{[Y^{\mathrm{T}}YF+\gamma(FYY^{\mathrm{T}})+\gamma F]_{ij}} \tag{3}$$

이로부터 무리짓기를 위한 알고리듬은 다음과 같다.

① 기초핵함수로 리용할 핵함수 $K_1,K_2,\cdots,K_s$들을 선택한다.

② 기초핵함수들을 리용하여 류사도행렬들을 계산하고 RCut를 리용하는 경우에는 방법 1로, NCut를 리용하는 경우에는 방법 2로 표준화한다.

이렇게 얻어진 류사도행렬들을 $W_1,W_2,\cdots,W_s$라고 하자.

③ α,Y,F를 우연적으로 초기화한다.

④ 반복회수에 이를 때까지 갱신규칙 (2), (3)을 반복한다.

⑤ $css_i=\arg\max_j Y_{ij}$라면 x_i는 css_i번째 무리에 소속시킨다.

론문에서 제기한 핵함수구성방법의 효과성을 확인하기 위하여 핵함수를 개별적으로 리용하는 경우와 여러개의 핵함수를 우에서 제기한 방법으로 결합하여 리용하는 경우의 무리짓기성능을 비교하였다.

실험에는 Dermatology, Glass, Soybean, Vehicle, Zoo라고 부르는 5개의 UCI자료를 리용한다.(표 1)

표 1. Dermatology, Glass, Soybean, Vehicle, Zoo자료

자료	표본수	표본벡토르의 차원수	무리수
Dermatology	366	33	6
Glass	214	9	6
Soybean	47	35	4
Vehicle	846	18	4
Zoo	101	16	7

알고리듬의 성능을 평가하기 위하여 선행연구[4]에서와 같이 무리짓기의 정확성을 평가하는 척도 $ACC=\left(\sum_{i=1}^{n}\delta(g_i,\ \mathrm{map}(c_i))\right)\Big/n$를 리용한다. 여기서 c_i는 무리짓기알고리듬을 적용했을 때 표본 x_i가 속하는 무리번호, g_i는 실지자료에서 x_i가 속하는 무리번호이다.

기초적인 핵함수로는 보통의 내적으로 정의된 핵과 가우스핵을 리용한다. 즉 $K_1(x, x') = x^{\mathrm{T}}x'$, $K_{2,\sigma}(x, x') = \exp(-\|x-x'\|^2/\sigma^2)$ 을 리용한다.

선행연구[4]의 결과와 비교하기 위하여 σ 는 선행연구[4]에서와 꼭같은 값으로 설정한다. 또한 $\mu = 100$, $\gamma = 10$ 으로 고정시키며 갱신회수는 300으로 정한다.

그림에서 보는바와 같이 갱신회수가 300이면 수렴성이 뚜렷이 나타난다.

무리짓기의 결과는 초기화에 의존하므로 선행연구[4]에서와 같이 독립적으로 256번 초기화하여 ACC 를 계산하고 평균하여 성능을 평가한다.

론문에서 제기한 핵함수구성방법으로 NCut를 리용했을 때와 RCut를 리용했을 때의 결과와 선행연구[4]의 결과를 보여주었다.(표 2, 3)

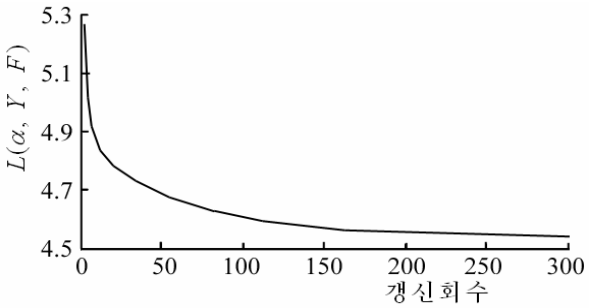

그림. Dermatology자료로 실험할 때의 갱신회수에 따르는 목적함수 $L(\alpha, Y, F)$ 의 변화

표 2. NCut를 리용한 경우의 결과

자료	보통의 내적/%	가우스 핵/%	결합핵 /%
Dermatology	74.7	87.5	89.4
Glass	46.3	50.2	54.3
Soybean	69.1	75.8	77.4
Vehicle	37.6	43.7	48.7
Zoo	64.7	80.3	81.4

표 3. RCut를 리용한 경우의 결과

자료	보통의 내적/%	가우스 핵/%	결합핵 /%
Dermatology	73.4	86.2	88.7
Glass	46.0	52.8	55.9
Soybean	73.1	78.5	84.7
Vehicle	39.3	43.1	51.4
Zoo	67.3	65.8	76.9

우의 결과로부터 론문에서 제기한 핵함수구성방법을 적용하면 개별적인 핵함수를 리용할 때보다 더 좋은 무리짓기결과를 얻는다는것을 알수 있다.

참 고 문 헌

[1] H. Zha et al.; Advances in Neural Information Processing Systems, 14, 105, 2002.

[2] H. Lu et al.; Pattern Recognition, 47, 418, 2014.

[3] R. Shang et al.; Pattern Recognition, 55, 172, 2016.

[4] D. Tolic et al.; Pattern Recognition, 82, 40, 2018.

주체109(2020)년 6월 5일 원고접수

A New Method to Construct Kernel Function for the Nonlinear Non-negative Spectral Clustering

Kim Jin Song, Kang Song Hyok

We propose a constructing method of kernel function appropriate to the given data and algorithm for solving the nonlinear non-negative spectral clustering using it and verify its effectiveness by experiments.

Keywords: spectral clustering, kernel function