The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Kernel Graph Laplacian Features による非線形な多様体の次元圧縮

Dimension reduction of nonlinear manifolds by Kernel Graph Laplacian Features

髙橋 春輝 Kazuki Takahashi　　竹川 高志 Takashi Takekawa

工学院大学情報学部
Faculty of Information Kogakuin University

Spectral clustering is used for clustering nonlinear manifolds. The better the performance, the better the distance between the clusters. However, as the S / N ratio of the data decreases, spectral clustering does not work well. This is because the pre-processing in spectral clustering is based on the assumption that the clusters can be sufficiently separated from each other by parameter adjustment. Therefore, in this paper, we propose robust preprocessing to S / N ratio by kernelized graph Laplacian features (Kernel GLF). The GLF is a linear transformation that brings each other's data with high affinity closer and keeps each other's data with low affinity away. The results show that Kernel GLF can convert nonlinear manifold into linear structure. By clustering the pre-processed data with K-Means, it became a more robust algorithm for S / N ratio than Spectral Clustering.

1. はじめに

クラスタリングでの解析では，対象となるデータの構造に結果は影響を受けることになる．データ構造が線形な場合には，K-Means[1]でクラスタリングを行い，各クラスタを分離することが可能である．一方，データ構造が非線形な場合にクラスタリングを行うには，データに前処理を施して線形分離可能な構造への変換が必要となる．そのようなアルゴリズムとして， Spectral Clustering[2][3]が知られている．

Spectral Clustering では各データ同士の類似度から，データ間の繋がりを表現し前処理を行う．この前処理では，各クラスタに属するデータ同士は直交させることが可能である（例として，クラスタ 1 のデータは(1, 0)の座標に，クラスタ 2 のデータは(0, 1)の座標に変換可能），ことを前提にしている．そのため前処理後のデータは K-Means で分離可能となる．しかしクラスタの SN 比が低くなると，異なるクラスタに属するデータ間の距離が近づき，同一クラスタに属するデータ間の距離は離れるようになる．そのような場合に Spectral Clustering を適用すると，過度に密集したクラスタができるような結果が現れる．これは前処理における前提を満たすためには，異なるクラスタに属するデータ同士が無限の距離をとることを必要とするために起こる．SN 比が小さな場合でもパラメータ設定により，異なるクラスタに属するデータ同士の距離を離すことができるが，それは同時に同一クラスタに属するデータ同士の距離も離してしまう．このことにより，SN 比の低いクラスタ同士を分離するような前処理は Spectral Clustering の前提では困難となる．

このことから，クラスタ間は有限の距離をとることを前提に前処理を施すことを考える．このような場合では，類似度の高いデータ同士は近づけ，類似度の低いデータ同士は遠ざけるように変換する方法が想定できる．このようなアルゴリズムには Graph Laplacian Features[4]がある．

Graph Laplacian Features(以降 GLF)は，類似度の高いデータ同士の変換後のユークリッド距離を近づけることと，変換後のデータ全体の分散を最大にすることを制約にしてデータの線形変換を行う．変換後のデータの分散を最大にすることは，類似度の低いデータ同士のユークリッド距離を遠ざけるように作用する．GLF はこの条件を満たす変換行列を固有値問題の解として得ることで，データを線形変換する．したがって，データ間の繋がりを考慮した前処理を可能とするが，元データの線形変換である以上，非線形な構造をとるデータから線形分離可能な構造への変換はできない．

そこで，元データの特徴量の線形変換とすることでカーネル化した GLF を提案し，データ間の繋がりから非線形な多様体の表現を可能とする前処理を試みる．そして前処理後のデータを K-Means で分類し，非線形かつ SN 比の低いデータに対するクラスタリングの堅牢性を実験的に検証する．

連絡先：竹川 高志，工学院大学 情報学部，〒163-8677 東京都新宿区西新宿 1-24-2，03-3340-0103，E-mail: takekawa@cc.kogakuin.ac.jp

2. Kernel GLF の定式化

2.1 GLF 概要

GLF は類似度行列

$$W_{ij} = e^{\left\{-\|x_i - x_j\|^2 / \sigma^2\right\}}$$

によって，変換後のデータ間に制約を与える．ここで現れる，σ はパラメータであり詳細は 2．3 節で述べる．また

$$D_{ij} = \begin{cases} \sum_k W_{ik} & i = j \\ 0 & i \neq j \end{cases} \qquad \mathrm{L} = \mathrm{D} - \mathrm{W}$$

$$P = diag(D_{11} \quad \cdots \quad D_{nn})/\mathrm{tr}(D)$$

を定義し，二つの制約の下最適化問題に定式化していく．

元データ $X_{m\times n} = \{x_1 \quad \cdots \quad x_n\}$ を変換行列 $A_{m\times d} = \{a_1 \quad \cdots \quad a_d\}$によって線形変換したデータを$Y_{d\times n} = A^T X$する．この時，一つ目の制約を

$$c_1(A) = \sum_i^n \sum_j^n \|y_i - y_j\|^2 W_{ij} = \sum_i^n \sum_j^n \|A^T x_i - A^T x_j\|^2 W_{ij}$$

と定義する．このときを$\min_A c_1(A)$満たす変換行列は類似度が高いほど距離が近づくように作用する

二つ目の制約を$\bar{y} = \sum_i^n y_i P_{ii}$, $\bar{x} = \sum_i^n x_i P_{ii}$として，

$$c_2(A) = \sum_i^n \|y_i - \bar{y}\|^2 P_{ii} = \sum_i^n \|A^T x_i - A^T \bar{x}\|^2 P_{ii}$$

と定義する．このとき$\max_A c_2(A)$を満たす変換行列は類似度が低い程距離を離すように作用する．

制約c_1，c_2の最適化を同時に行うために，損失関数を

$$\mathrm{Cost}(A) = {}^{c_1}/_{c_2}$$

と定義し，$\max_A Cost(A)$の解として変換行列$A_{m\times d}$を求める．

本文

2.2 GLF のカーネル化

データ$X_{m\times n}$の特徴量ベクトルを$\Phi_{r\times n} = \{\phi_1 \quad \cdots \quad \phi_n\}$として，$\bar{\phi} = \sum_i^n \phi_i P_{ii}$，$\tilde{\Phi}_{r\times n} = \Phi - \bar{\phi}$と定義する．この時変換後のデータを$Y_{d\times n} = A_{r\times d}^T(\Phi - \bar{\phi}) = A^T\tilde{\Phi}$とする．そして$A$を，$\tilde{\Phi}$の線形和の成分$B_{n\times d} = \{b_1 \quad \cdots \quad b_d\}$と，$\tilde{\Phi}$に直交する成分$\Xi_{r\times d}$に分割する．するとカーネル関数$K_{ij} = \langle\phi_i, \phi_j\rangle$を用いて，$\tilde{K}_{ij} = \tilde{\phi}_j^T\tilde{\phi}_i = K_{ij} - \sum_s^n K_{is}P_{ss} - \sum_t^n K_{jt}P_{tt} + \sum_s^n\sum_t^n K_{st}P_{ss}P_{tt}$ とし，$Y_{d\times n} = B_{n\times d}^T\tilde{K}_{n\times n}$を得る．

このとき二つの制約は，$c_1(B) = 2\sum_k^d b_k^T\tilde{K}L\tilde{K}b_k$，$c_2(\boldsymbol{B}) = \sum_k^d b_k^T\tilde{K}P\tilde{K}b_k$となるため，全体の損失関数は，

$$\mathrm{Cost}(B) = \sum_k^d {b_k^T\tilde{K}P\tilde{K}b_k}\Big/{b_k^T\tilde{K}L\tilde{K}b_k} \quad ①$$

となる．ここでラグランジュ未定乗数法により，

$$\mathrm{L}(b_k,\ \lambda_k) = b_k^T\tilde{K}P\tilde{K}b_k - \lambda_k(b_k^T\tilde{K}L\tilde{K}b_k - 1)$$

$$\frac{\partial \mathrm{L}(b_k,\ \lambda_k)}{\partial b_k} = \tilde{K}P\tilde{K}b_k - \lambda_k\tilde{K}L\tilde{K}b_k = 0$$

$$\tilde{K}P\tilde{K}b_k = \lambda_k\tilde{K}L\tilde{K}b_k \quad ②$$

となり，d個の固有ベクトルbが$B_{n\times d}$の要素となる．

2.3 パラメータについて

Kernel GLF に現れるパラメータは 3 つとなる．

類似度のパラメータ σ を決める方法は，先行論文[1]を参照した．その方法は，データの近傍数Kをパラメータとして与え，すべてのデータで各データからK番目に遠いデータ点とのユークリッド距離を求め，それらの平均を局所的なデータの広がりとして基準 σ とする．基準よりも，データ間が離れているか否かを指標に類似度を決める方法となっている．

変換後の次元数dは，①②より全体の損失関数が

$$\mathrm{Cost}(B) = \sum_k^d \frac{b_k^T\tilde{K}P\tilde{K}b_k}{b_k^T\tilde{K}L\tilde{K}b_k} = \sum_k^d \frac{\lambda_k b_k^T\tilde{K}L\tilde{K}b_k}{b_k^T\tilde{K}L\tilde{K}b_k} = \sum_k^d \lambda_k$$

となるため，d個の固有値の和と全固有値の和との割合が，与えた寄与率より大きくなるように dを定める．

カーネル関数のパラメータの設定は手動で行う．また，カーネル関数自体の設定も任意である．

3. Kernel GLF の検証

以下の検証では，カーネル関数は RBF カーネルを用いた．

$$\text{RBF カーネル}\quad k_{ij} = e^{\left\{-\|x_i-x_j\|^2/\gamma\right\}}$$

3.1 Circle データの Kernel GLF による前処理

sklearn のデータセットの一つである circle データに対して，ガウスノイズが 0. 05，0. 1，0. 15，0. 2 のそれぞれに Kernel GLF を施した結果を図 1 に示す．

近傍数 K は 100，削減後の次元数dは 2，γは 50 とした．

図 1 から，Kernel GLF により類似度に基づき線形分離可能な構造に変換できていることがわかる．

3.2 Circle データのクラスタリング結果

データ数 1000 の circle データを，ノイズが 0. 05 から 0. 2 まで 0. 01 刻みで各 100 回データを作り直し，正答数の平均と標準偏差を求めた．図 2 は，横軸がノイズ，縦軸が正答数を示す．

近傍数 K は 100，寄与率は 0. 95，γは 5 から 100 までの 5 の倍数とした．また，Kernel PCA では RBF カーネルは

$$\text{RBF カーネル}\quad k_{ij} = e^{\{-\gamma\|x_i-x_j\|^2}$$

となっている．

図 2 より，Spectral Clustering での正答数はノイズが 0.1 以上になると，急激に下がり始めるのがわかる．一方，Kernel GLF 後に K-Means を行った正答数は横ばいである．Kernel GLF と Kernel PCA ではカーネル関数が同じではあるが，K-Means でのクラスタリング結果である正答数には大きな差がみられる．

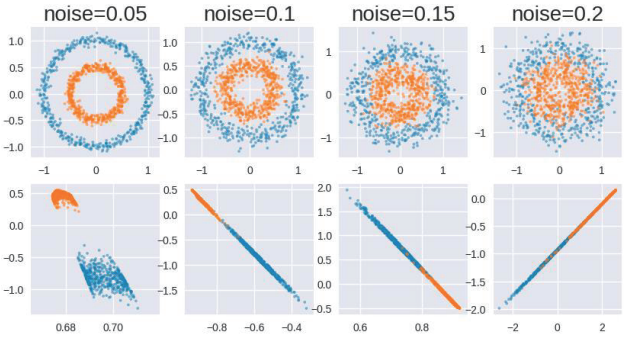

図 1:noise 毎の circle データ分布(上段). Kernel GLF 後のデータ分布(下段)

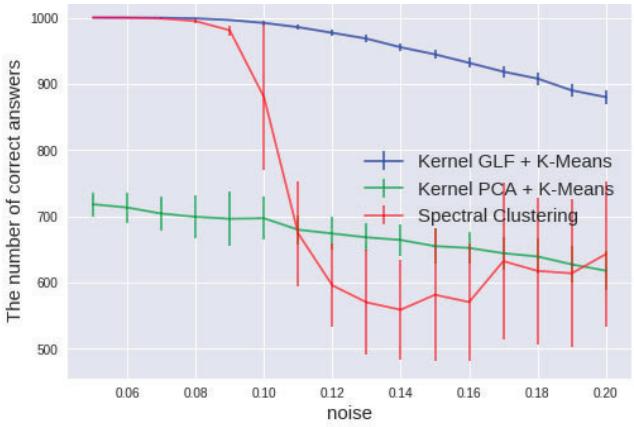

図 2:アルゴリズム毎の正答数の平均と標準偏差

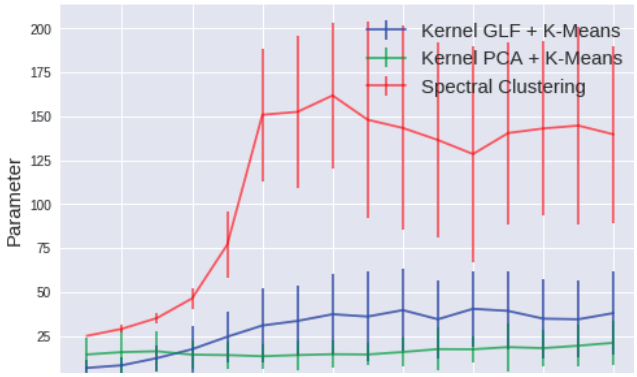

図 3:アルゴリズム毎のパラメータの平均と標準偏差(Kernel GLF と Kernel GLF, Kernel PCA でのパラメータはカーネル関数のγを，Spectral Clustering のパラメータは類似度の $\frac{1}{\sigma^2}$を示している)

4. まとめと考察

3 章の検証では, Kernel GLF により, 非線形な構造が線形な構造に変換されることが確認された. このことから, 変換後の分散だけでなく類似度を考慮することで, Kernel PCA での前処理よりも非線形な多様体を表現しようとする作用が現れることが, Kernel GLF では期待できる. しかし今回の検証は, 非線形な多様体構造であるものの比較的単純なデータであるため, 高次元となるような複雑なデータである場合でも検証が必要となる. しかしながら, カーネル関数を使用する以上, Kernel GLF により得られる結果は恣意的にならざるを得ない. Kernel PCA との比較でも見て取れるように, カーネル関数が等しくても GLF と PCA のアルゴリズムの根本的な違いにより挙動は変動する. したがって, Kernel GLF におけるカーネル関数とパラメータの選択をする際の, 客観的判断に基づいた指標を定めることが今後の課題となるだろう.

参考文献

[1] J. B. MacQueen. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathmatical Statistics and Probability, volum 1, pages 281-297.

[2] Ulrike von Luxburg. (2007). A Tutorial on Spectral Clustering. Max Planck Institute for Biological Cybernetics.

[3] MalikShi and JitendraJianbo. (2000). Normalized Cuts and Image Segmentation. IEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8.

[4] SpenceGhanbari Panos E. Papamichalis and LarryYasser. (2010). Graph-Laplacian Features for Neural Waveform Classification. IEEE transactions on bio-medical engineering, VOL. 58, NO. 5.