位相多様体学習方式の提案
Topological Manifold Learning

田﨑 元・システム分科会・中央大学大学院

Dimension reduction is one of the most important issues in machine learning and computational intelligence. Typical data sets are point clouds in a high dimensional space with a hidden structure to be found in low dimensional sub manifolds. Finding this intrinsic manifold structure is very important in the understanding of the data and for reducing computational complexity. In this paper, we propose a novel approach for dimension estimation of topological manifolds based on measures of simplexes. We also investigate the effects of resolution changes for dimension estimation in the framework of Morse theory. The result is a method that can be used for data located in simplical complexes of varying dimensions and with no continuous or differentiable structure.

1. Introduction

機械学習や計算知能の研究において，次元削減は重要な問題のひとつである．多くのデータセットでは，高次元空間におけるデータ点群が実質的に低い次元の部分多様体構造を持つことが多い．

本研究では，位相多様体に対する次元推定手法を提案する．さらに，Morse理論やスケールスペースに基づいて，次元推定に対する位相的なフレームワークを検証する．

2. Proposal Method

■ Mathematical Definition

➢ n-単体の測度 V_n ：

$n = 1, 2, 3$ における長さや面積，体積を一般化した概念で，単体の大きさを表す．n-単体から$(n-1)$-単体への直交射影の距離h_kを用いて，次式で計算することができる．

$$V_n = \frac{1}{n!}\prod_{k=1}^{n} h_k \,.$$

➢ 正規化測度（Measure-ratio）：

単体測度V_n を1-単体の長さV_1 の平均で正規化する．
($\bar{V}_i$ ：i-単体の測度の平均)

$$M_n = \frac{\bar{V}_n}{\bar{V}_1{}^n/n!} = \prod_{k=1}^{n}\frac{m_k}{m_1}$$

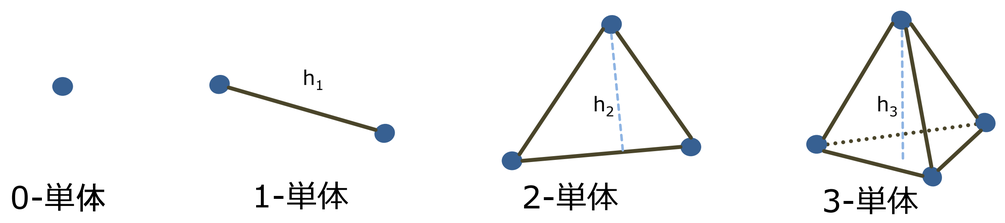


Fig.1 単体

■ Algorithm

入力: $X = (\boldsymbol{x}_i)$, $\boldsymbol{x}_i = [x_1, \dots, x_D]$ $(i = 1, \dots, N)$

1. データ点群Xの近傍分割を行う．
2. 各近傍において，次の2ステップにしたがって，単体の構築と単体測度の計算を行う:
   a. 単体測度V_1が最大となるよう1-単体を作成する．
   b. $n \geq 2$ に対して，直交射影の長さh_nが最大となるような点を探索．n-単体を決定して，単体測度V_nを計算する．
3. 各近傍で正規化測度M_nを計算し，次元ごとに比較を行う．

多様体がd次元のとき，正規化測度M_nは$n < d$であれば，M_nは安定して1に近い値をとる．一方で，$n > d$においては，M_nの値は急激に小さくなる．この傾向を次元推定の基準として推定を行う．

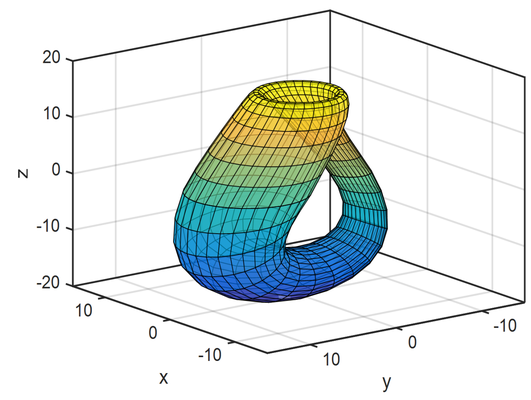
3. Simulation

本研究では，提案手法を局所主成分分析(LPCA)との比較により検証を行う．各手法において，近傍分割には一般化Lloydアルゴリズムを用いて，次元推定を行う近傍に分割する．

データセットは，Klein Bottleから20000点をランダムにサンプリングすることで作成した．

■ Klein Bottle

$$\begin{aligned} x_1 &= (R + r\cos\phi)\cos\theta \\ x_2 &= (R + r\cos\phi)\sin\theta \\ x_3 &= r\cos\left(\frac{\theta}{2}\right)\sin\phi \\ x_4 &= r\sin\left(\frac{\theta}{2}\right)\sin\phi \end{aligned}$$



■ 推定結果

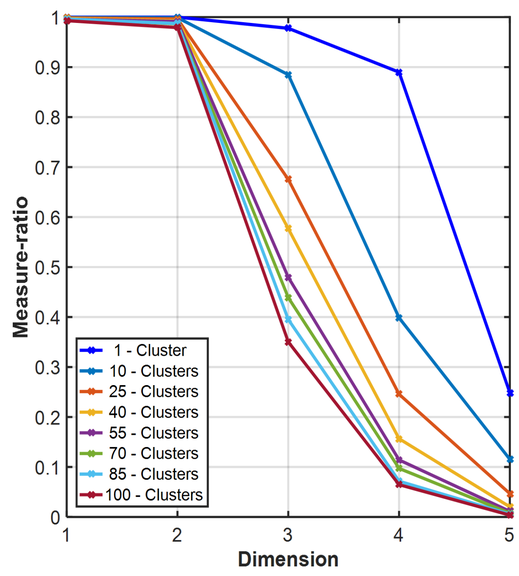


Fig.2 単体測度の推移

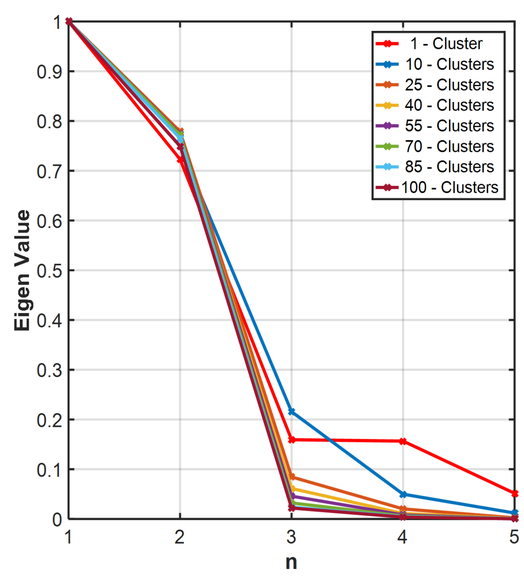


Fig.3 固有値の推移

4. Conclusion

本研究では，単体測度に基づく次元推定法を提案し，従来法であるLPCAとの比較により，提案手法が有効にはたらく場合を確認した．さらに，Morse理論に基づく異なる近傍サイズにおける次元推定では，位相的に安定な推定が行えることを確認した．データ多様体の位相構造の抽出や低次元空間への埋め込み等が，今後の課題である．

References

1. Y. Ma and Y. Fu, *Manifold learning theory and applications,CRC Press, 2011.*
2. N. Kambhatla and T. K. Leen, “Dimension reduction by loca l principal component analysis”, *Neural computation*, vol.9, no. 7, pp. 1493 - 1516, 1997.