

論文メモ

文献番号	0010
日付	2021 年 12 月 02 日
名前	武川海斗

文献情報

著者	Yukihiro Hamasuna, Yasunori Endo
英文タイトル	On Kernelized Sequential Hard Clustering
和文タイトル	逐次抽出型ハードクラスタリングのカーネル化
書誌情報	Joint 8th International Conference on Soft Computing and ,pp. 416–419 (ページ), 2016
キーワード	sequential cluster extraction, sequential hard c- means, noise clustering, kernel function.

1 論文の要約

本論文は、逐次抽出法をカーネル法に組み込んだ手法について述べた論文である。カーネル c -means を基にして、逐次抽出法への拡張を検討している。実験としては、非線形なデータセットに対して有効なクラスタリングが行えるかを確かめている。しかし、実験結果は否定的なものであり、非線形なデータ群に対しては有効なクラスタリング結果を得ることができなかった。今後の課題としては、他のカーネル関数やハイパーパラメータを広く実験することが必要であると述べられている。

2 論文のトピック

本論文では、逐次抽出法をカーネル法を用いたクラスタリングに導入することを目的としている。ここでは、カーネル c -means を対象として実験を行っている。

3 ベースとなった手法

3.1 逐次抽出型 c -means(Sequential Hard c -Means; SHCM)

ハード c -means に逐次抽出型を拡張した数式が以下の式 (1) であり、制約条件は、式 (2) である。 $D > 0$ はノイズパラメータであり、自分で与える必要がある。ここで、 u_{k1} は抽出クラスタに属する帰属度、 u_{k0} はノイズクラスタに属する帰属度を表している。データ点とクラスタ中心のノルム 2 乗値と D を比較して、どのクラスタに含まれるか判断する。

$$J_{sh}(U, V) = \sum_{k=1}^n u_{k1} \|x_k - v_1\|^2 + \sum_{k=1}^n u_{k0} D \quad (1)$$

$$\mathcal{U}_{sh} = \left\{ (u_{ki}) : u_{ki} \in \{0, 1\}, \sum_{i=0}^1 u_{ki} = 1, \forall k \right\} \quad (2)$$

3.2 ν -逐次抽出型 c -means(ν -Sequential Hard c -Means; ν SHCM)

SHCM における D の定義を変えた手法で、いわば SHCM の亜種と言える手法である。ここで、 nu はノイズクラスタの個数を表している。つまり、自分でノイズクラスタの個数をハイパーパラメータとして与える必要がある。そのため、この手法はクラスタの構造を考慮することができず、非線形なデータ構造などには適していない。

$$J_{\nu sh}(U, V, D) = \sum_{k=1}^n u_{k1} \|x_k - v_1\|^2 + \sum_{k=1}^n u_{k0} D - \nu D \quad (3)$$

4 提案手法のコア要素

SHCM に対して、カーネル法を導入した手法である。ここで、カーネル法を導入することで、非線形データに対してのクラスタリングを行うことを目的とする。

$$J_{ksh}(U, W) = \sum_{k=1}^n u_{k1} \|\phi(x_k) - W_1\|^2 + \sum_{k=1}^n u_{k0} D \quad (4)$$

5 実験デザイン・結果と考察

KSHCM の性能を評価するために、非線形なクラスタ構造を持つ、データ数 150 個、クラスタ数 2 の人工データを用いて評価を行った。ハイパーパラメータ D, β を 3 つの組み合わせで図にプロットしている。しかし、結果は否定的なものであり、非線形な構造を抽出することはできなかった。考察として、適切なカーネル関数やハイパーパラメータを広く実験することが必要であると述べている。ここで、興味深いのは、カーネル主成分分析を用いたクラスタ構造の可視化について必要であると述べていることである。この考えを自分の研究に組み込んでいくつもりである。

6 手法の限界・今後の課題

この論文で提示された実験だけでは、手法について良好な結果が得られなかった。今後、良好な結果を出すためには、他のカーネル関数やハイパーパラメータで実験することが必要であると考え。また、主成分分析を用いたクラスタ構造決定も今後研究の材料になる。

7 特に重要な関連研究

次に読むべき論文のリスト

[1] の論文は、カーネル法を導入することによって、線形分離可能性が高まることで、関連するデータ群が単純化されるという仕組みを扱っている。これにより、クラスタ数を推定することを目的とした論文である。Girolami M., Mercer kernel-based clustering in feature space, IEEE Transactions Neural networks, Vol. 13, No. 3, pp. 780—pp. 784, 2002.