

論文メモ

文献番号	0001
日付	YYYY 年 MM 月 NN 日
名前	情報太郎

文献情報

著者	春山秀幸, 遠藤靖典, 大久保貴義
英文タイトル	
和文タイトル	カーネル関数を用いた階層的クラスタリング
書誌情報	知能と情報 (日本知能情報ファジィ学会誌), Vol. 17, No. 4, pp. 459–467, 2005
キーワード	階層的クラスタリング, カーネル関数, パターン認識

1 論文の要約

本論文では, 階層型クラスタリングに対して, カーネル関数を導入する手法を提案を行っている. 結果として, 非線形なデータ構造を持つクラスタ構造に対して, 有効にクラスタリングを行うことができるようになった. 既存の階層型クラスタリング手法と, 非階層型クラスタリング手法を比較し, 人工データと実データを用いて実験を行った.

2 ベースとなった手法

2.1 階層型クラスタリング

階層型クラスタリングにおけるクラスタリングの流れは, 各データ間に非類似度を定義し, この非類似度に基づいてクラスタを逐次結合していくというものである. ここで, 階層的クラスタリングには階層併合的方法と階層分割的方法があるが, 本論文では階層併合的方法を用いる. 以下にアルゴリズムを記す.

AHC アルゴリズム データの集合を $X = x_1, \dots, x_n$, 非類似度を $d(x, y)$ とする

STEP1 初期設定として, n 個のデータを n 個のクラスタとする.

$$G_i := \{x_i\} \quad (1)$$

$$d(G_i, G_j) := d(x_i, x_j) \quad (i, j = 1, \dots, n) \quad (2)$$

STEP2 クラスタ間の非類似度行列 $D = d[d_{ij}]$ を更新する.

STEP3 D を参照し, 最も類似性の高い 2 つのクラスタを結合して一つのクラスタを生成する.

$$s(G_q, G_r) = \max_{i,j} s(G_i, G_j) \quad (3)$$

$$d(G_q, G_r) = \min_{i,j} d(G_i, G_j) \quad (4)$$

STEP4 クラスタ数が 1 なら終了. そうでないなら, **STEP2** に戻る.

3 提案手法のコア要素

本論文の提案手法は, 階層併合的手法にカーネル関数を適応した, K-AHC(Kernel-agglomerative hierarchical clustering) の提案を行っている.

3.1 K-AHC クラスタリングアルゴリズム

K-AHC において非類似度 $d_\phi(x, y) = d(\phi(x), \phi(y))$ は特徴空間における距離の自乗 $\|\phi(x) - \phi(y)\|^2$ とする．特徴空間での x_i と x_j との距離の自乗を，カーネル関数を用いることで ϕ の形が分からなくても計算することができる．あとは，既存のアルゴリズムにこの非類似度を導入すれば良い．

4 実験デザイン・結果と考察

実験では、人工データと Iris データを用いている．提案手法である K-AHC との比較として、カーネル関数を用いない従来の階層型クラスタリングと非階層型クラスタリングを使用した．非階層型クラスタリングとしてファジィ c-平均法を用いた．階層型クラスタリングとして、最短距離法、最長距離法、群間平均法、群内平均法、重心法、Ward 法を用いた．また、カーネル関数として多項式カーネル、ガウス型カーネル、パーセプトロンカーネル、シグモイドカーネルを用いた．データによっては、従来の手法より良好な分類結果が得られた．また、特定のデータに対して有効なカーネルを確認できた．

階層型クラスタリングとカーネル関数に相性があると考えらる．具体的には、内積を扱うカーネル関数か、ノルムを扱うカーネルかどうかである．実験では、Ward 法とノルムを扱うガウス型カーネルと相性が良かった．Ward 法はクラス間やクラス内の類似度を計算するため、ガウス型カーネルは原点の影響を受けないことが理由として考えられる．

5 手法の限界・今後の課題

これはカーネル法の課題と言えるのだが、高次元空間でデータがどのように分割しているかを把握することが困難な問題がある．これは写像関数 ϕ がように現れないためであり、高次元の分布状況を把握することは今後の課題となる．

6 特に重要な関連研究

参考文献 [1] は、ファジィ理論が初めて提唱された論文であり、学術的にも価値のある論文である．今後、GPSCRM のファジィ化を進めるにあたって、読むべきである論文であると言える．

次に読むべき論文のリスト

- [1] L. A. Zadeh, Fuzzy sets, *Information and Control*, vol. 8, pp. 338–pp. 353, 1965