

機関番号：12102

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500252

研究課題名（和文） 高次 Aggregation Operator の開発とクラスタリングモデルへの適用

研究課題名（英文） Development of a high dimensional aggregation operator and its application to a clustering model

研究代表者

イリチュ 美佳（佐藤 美佳）（SATO-ILIC MIKA）

筑波大学・大学院システム情報工学研究科・准教授

研究者番号：60269214

研究成果の概要（和文）：高次元Aggregation Operatorを提案し、このOperatorの各種特性を精査するとともに、クラスタリングモデルへ適用し、実用化を図った。従来、統計的測度空間に基づく研究がなされてきたが、より一般的に表現した高次元Aggregation Operatorは、未だ開発されていなかった。そこで、高次元Aggregation Operatorを開発し、より汎用性の高いクラスタリングモデルを実現した。

研究成果の概要（英文）：This research proposed a high dimensional aggregation operator and investigated its features. In addition, we applied this aggregation operator to a clustering model for practical use. Conventionally, research based on statistical metric space has been discussed for this area, however a generalized high dimensional aggregation operator has not been proposed. So, in this research, we developed a high dimensional aggregation operator to obtain a general purpose clustering model by exploiting this aggregation operator.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,500,000	450,000	1,950,000
2009年度	1,200,000	360,000	1,560,000
2010年度	700,000	210,000	910,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：分類、パターン認識、Aggregation Operator

1. 研究開始当初の背景

クラスター分析は、分類という操作を通じてデータに内在する共通性や類似性を探索

し、その特徴をまとめデータの構造や特性を抽出するデータ解析法である。そのため、データ間の類似性を如何に定義するかが重要な問題である。

統計・確率の分野では、二つの個体間の距離が確率構造を持つものと仮定した統計的測度空間（確率的測度空間）に基づく T-norm が定義され、数学的見地で、研究がなされてきた。近年では、逆に、具体的データを基にして可能な限り、データの性質を保存するように類似性を定義し、どのような数学的性質を包含させるべきかが研究されている。このような類似度を示す計量が完成すれば、これを用いたクラスター分析の結果が、より具体的な意味をもつものと考えられる。この研究は、T-norm をより一般的に表現した Aggregation Operator の研究として広く知られている。しかし、高次元直積空間上の Aggregation Operator については、未だ開発されていない。そこで、本研究では、高次元直積空間における Aggregation Operator を開発しようとするものであり、その位置づけは、高次元直積空間における新たなメトリックの提案と、クラスターリングモデルによる性能評価である。

2. 研究の目的

本研究の目的は、高次元 Aggregation Operator を提案し、その性質を解明し、クラスターリングモデルに導入することにより、その有効性を検証することである。研究代表者は、これまでクラスター分析の手法の一つとして、類似度の構造モデルを仮定し、その構造を推定することにより、データを分類するファジィクラスターリングモデルを提案してきた。また、類似度の構造を Aggregation Operator を用いてより一般的に表現し、これを用いた加法的モデルの性質を調べ、その妥当性および有効性について検討を行ってきた。しかし、これらのクラスターリングモデルでは、現在の複雑なデータの構造を、論理的客観性を伴う形で要約し、説明することが難しい。その原因の一つは、データの分類をデータの変数からなる個体空間に限定していることにあると考えられる。そこで、より高次元の識別空間で分類を行うことが必要である。そのためには、まず、この空間で使用できる何らかの計量が必要である。そこで、Aggregation Operator をこの空間で定義し、これをクラスターリングモデルに導入することにより、より汎用性の高いモデルを開発し、その有効性を検証することが本研究の目的である。

3. 研究の方法

高次元空間における Aggregation Operator

を定義した。その性質を、推定するパラメータ空間の次元による相違性と、クラスター間の交互作用量を考慮した場合のモデルの非線形性の観点から調査した。その際、加法モデルに、カーネル法を利用し、高次元ヒルベルト空間におけるパラメータ推定で、分類構造を抽出するクラスターリングモデルを提案した。クラスターリングモデルへ、提案した Aggregation Operator を適用し実データに当てはめることにより、評価を行い、随時、公表を行った。

また、高次元直積空間における一般化 Aggregation Operator を定義し、クラスターリングモデルに適用した。さらに、一般化クラスターリングモデルの性質と、従来のクラスターリングモデルの関連性を解析した。応用として、ノイズが多くデータの明確な構造抽出が困難であるとされる脳波データに適用し、人間の思考判別に対して、応用可能であることを示した。さらに、大規模データへの適用可能性を検証するため、時系列画像データを用いて、応用可能性を調査した。その際、高次元データの次元縮約に当たり、主成分分析に高次元の距離構造を導入することで、新たな手法を開発した。開発した高次元空間における Aggregation Operator の実データへの応用について、その動作確認を行い、一般化 Aggregation Operator の性質を精査した。また、Aggregation Operator を自己組織的に学習させる自己組織化加法的クラスターリングモデルへ適用する新たな方法論への拡張を行った。また、区間評価データに対する高次元 Aggregation Operator を提案手法に当てはめることにより、高次元のデータの分類構造をも考慮できるように拡張した。

4. 研究成果

(1) 研究の主な成果

①高次元空間における Aggregation Operator の定義とクラスターリングモデルの開発、及び適用

定義した Operator を、高次元ヒルベルト空間におけるパラメータ推定により、分類構造を抽出するクラスターリングモデルに導入した。このモデルは、加法モデルに、カーネル法を利用したものである。従って、パラメータ空間の次元による相違性と、クラスター間の交互作用量を考慮した場合のモデルの非線形性の差異から新たな Operator の性質の調査が出来る。調査の結果、ノイズが多いとされるデータに対して、有効な結果を得た。

②高次直積空間における一般化 Aggregation Operator の定義とクラスタリングモデルの開発、及び適用

この Operator を、クラスタリングモデルに適用し、一般化クラスタリングモデルの性質と、従来のクラスタリングモデルの関連性を解析した。応用として、ノイズが多くデータの明確な構造抽出が困難であるとされる脳波データに適用し、人間の思考判別に対して、応用可能であることを示した。さらに、大規模データへの適用可能性を検証するため、時系列画像データを用いて、応用可能性を調査した。その結果、計算時間を縮小し、より識別力の高い結果を得ることができた。その際、高次元データの次元縮約に当たり、主成分分析に高次元の距離構造を導入することで、新たな手法を開発した。

(2) 得られた成果の国内外におけるインパクト

これらの成果を、論文誌に公表し、さらに、招待講演を含む国内外の国際会議、学会発表で公表した。その際、下記のいくつかの学術的賞を受賞した。

・ 1st Runner up Award (theoretical development in computational intelligence), St. Louis (USA), 2010

・ Outstanding Contribution Award, Seventh International Symposium on Management Engineering, (keynote speech), 2010

・ Excellent Paper Award, The 6th International Symposium on Management Engineering, Dalian (China), 2009

・ Best Student Presentation Award, 5th International Workshop on Computational Intelligence & Applications, 2009 (指導学生の受賞)

(3) 今後の展望

Aggregation Operator の自己組織化構造を加法的クラスタリングモデルへ適用する拡張や、より汎用性の高いシンボリックデータへの応用、さらに高次元小標本データへの適用が考えられ、より複雑で大量なデータへの理論的方法論の拡張と応用が期待される。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 18 件)

- ① M. Sato-Ilic, Generalized Aggregation Operator Based Nonlinear Fuzzy Clustering Model, Intelligent Engineering Systems through Artificial Neural Networks, 査読有, Vol. 20, 2010, pp. 493-500 (1st Runner up Award (theoretical development in computational intelligence) 受賞)
- ② M. Sato-Ilic and D. Wu, Fuzzy Cluster Number Selection Based on Alignment of Similarities, The 6th International Symposium on Management Engineering, 査読有, 2009 (Excellent Paper Award 受賞)

[学会発表] (計 21 件)

- ① M. Sato-Ilic, Fuzzy Clusterwise Analyses in High Dimensional Space, Seventh International Symposium on Management Engineering (基調講演), 2010年8月27日, 北九州国際会議場, 福岡県小倉市 (Outstanding Contribution Award 受賞)
- ② M. Sato-Ilic, Fuzzy Classification Structure Based Analyses through High Dimensional Data Space, Joint meeting of Japan-Korea Special Conference of Statistics (招待講演), 2010年2月2日, 岡山大学, 岡山県岡山市

[図書] (計 1 件)

L.C. Jain, M. Sato-Ilic, M. Virvou, G.A. Tsihrintzis, V.E. Balas, and C. Abeynayake eds., Springer-Verlag (Germany), Computational Intelligence Paradigms, 2008, 195-217

6. 研究組織

(1) 研究代表者

イリチュ 美佳 (佐藤 美佳) (SATO-ILIC MIKA)

筑波大学・大学院システム情報工学研究科・准教授

研究者番号 : 60269214

(2) 連携研究者

清水 信夫 (SHIMIZU NOBUO)
統計数理研究所・助教
研究者番号：00332130