4K-1

サンプルの隣接関係に着目した多次元データに内在する共変関係検出に関する考察

名児耶　厚[†]　　石川　慎也[†]　　小野　裕次郎[‡]　　市野　学[†]
東京電機大学理工学部情報社会学科[†]　　十文字学園女子大学社会情報学部[‡]

1　はじめに

データ解析において，データに内在する特徴間の共変関係の検出は重要である．ピアソンの積率相関係数[1]は、特徴間の線形構造の評価を目的としている．非線形の共変関係を含むより広い共変関係の評価法として、カルホーン相関係数[2]が提案されている．しかし、これらの方法は、基本的に2特徴間の関係に注目している．

本研究では，2特徴の共変関係ばかりでなく、3特徴の共変関係も検出できる方法の構築を目的としている．特徴間に共変関係が存在する場合，ある特徴において近隣にあるサンプル対は，他の特徴においても近隣に存在する傾向にある(図1,2参照)．このような性質は、「幾何学的に薄い」サンプルの分布に共通しており,カルホーン相関係数においても利用されている．また、この性質は3特徴においても同様に保たれる．

本報告では，サンプル同士の隣接関係を複数の特徴で同時に調べることで，幾何学的に薄い構造を検出する手法について考察する．まず，2特徴で非線形の共変関係を有するデータを対象にした結果を示す．つぎに，その過程において取得できるパラメータが，2特徴，3特徴の「幾何学的に薄い構造」を有するデータに共通する，有用な情報となることを述べる．

2　共変関係の検出

2．1　隣接関係の定義

与えられたデータが，N個のサンプル ω_k (k=1,2,…,N)，d個の特徴 F_k(k=1,2,…,d)によって記述されているとする．p番目の特徴 F_p において，任意のサンプル対 ω_i,ω_j によって形成される閉区間を想定する．その閉区間に包含される他のサンプルが存在しないとき，これらのサンプル対 ω_i, ω_j は，特徴 F_p に関して相対近隣(Relative Neighborhood, RN)であるという[3]．

Detection of covariant relation in multi-dimensional data by using relative neighbors
†Atsushi Nagoya, Shinya Ishikawa, Manabu Ichino,
Department of Information and Arts, Tokyo Denki University
‡Yujiro Ono,
Department of Social Sciences and Information Science, Jumonji University

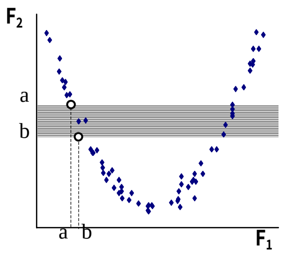

図1　幾何学的に薄い構造

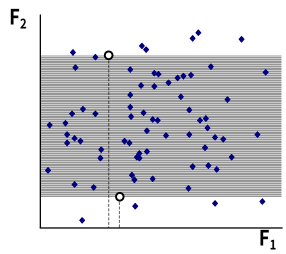

図2　幾何学的に厚い構造

次に，1特徴におけるRNが複数の特徴において同時に成り立っているサンプル間の関係を，MRN（Multi-feature Relative Neighborhood）とよぶ．MRNは，線形構造を含む単調な構造において，隣接サンプル対が満たす特有の性質である．

2．2　多価性への対処

特徴間に非線形構造を有するデータにおいては，その多価性によりMRNを満たさないサンプル対が現れる（図1参照）．さらにノイズの追加によっても，同様にMRNの成立が妨げられる．そこで，「RNが複数の特徴で同時に成立すること」としていたMRNの成立条件に対し，「対象となるサンプル対の間に含まれるサンプルの数がp個以下であることが同時に成立すること」という緩和条件を追加する．このときのpを許容度と呼び，許容度pの下にMRNとして成立するサンプル間の関係をMRN(p)と表すことにする．幾何学的に薄い構造を有するデータの場合,MRN(p)は小さな範囲の許容度pで成立し，逆に多次元の乱数構造など，幾何学的に厚い構造に対しては許容度pの増大に対して,MRN(p)の総数は緩やかな増加を見せる．

3　性能評価実験と検討

3．1　データ構造の骨格的表現

pを0から逐次増加していき，MRN(p)の成立したサンプル対を辺で結んで無向グラフとして表現する．ただし，グラフの生成において「連結グラフになったら終了」，という共通の終了条件を適用した．

図3に，2特徴間で非線形構造を有するデータ

に対して生成されたグラフを示す．ただし，見易さを優先するため，サンプル数は 100 とした．

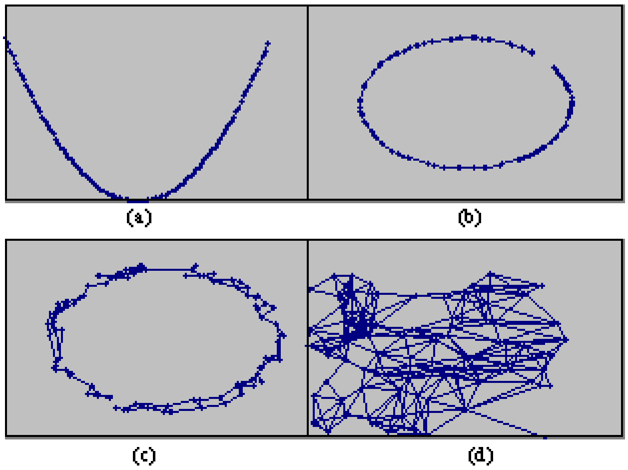

図 3　MRN(p)によるデータの骨格表現

図 3(a)は二次関数構造，(b)は円構造である．双方ともデータの骨格構造を表すグラフが生成されている．また，(a)(b)ともにノイズは加えていないが，前者はピアソンの積率相関係数 [1]，後者はカルホーン相関係数 [2]によって検出されない典型的な例である．また，図 3(c)は(b)に対してわずかなノイズを追加したデータである．元の円構造に対してノイズに相当する分の辺が追加されているが，データの骨格構造は保存されている．さらに，いずれも乱数構造である図 3(d)と比較し，データの共変関係を「幾何学的に薄い骨格」として捉えている．

実験の過程においては，線形構造や三角関数など他の共変関係を有するデータを用い，ノイズの割合を変えるなど，報告内容以外の試みも行っている．

３．２　許容度の分布を利用した評価式の検討

前節において, MRN(p)のグラフがデータの骨格構造を捉えられることを述べた．次に，「特徴間に共変関係が見られる場合，狭い範囲の許容度 p で MRN(p)が成立する傾向があること」に着目し，2 特徴，3 特徴それぞれのデータについて，乱数構造を有するデータとの比較実験を行った．図 4, 5 は，横軸に許容度，縦軸にその許容度で成立している総 MRN(p)数を表したグラフである．ただし，サンプル数は 250 である．また，3 特徴のデータは 2 特徴のデータに新たに 1 特徴を追加する方法で生成した．乱数構造以外は，3 次元空間中で，サンプル群がロープ状に並んだ形状となっている．

2 特徴，3 特徴いずれの場合においても，乱数構造と比較すると，狭い範囲の許容度において急激な動きが見られる．また，ノイズを付加すると，ノイズの度合いに応じて乱数構造へと近づいていく，という変化をすることも確認できており，これらを用いて評価式の作成が可能であると考えている．

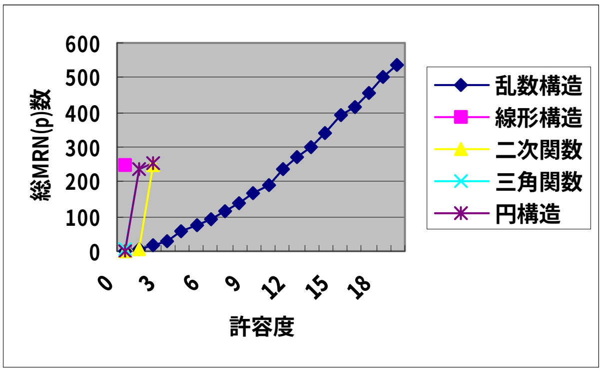

図 4　2 特徴の場合

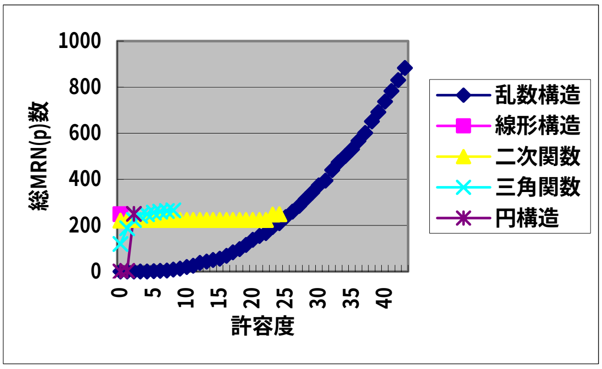

図 5　3 特徴の場合

4　おわりに

サンプルの隣接関係をもとに, MRN(p)という関係を定義し，幾何学的に薄い構造を 2 特徴，3 特徴において検出することを試みた．実験により，幾何学的に薄い構造に対して, MRN(p)は特徴的な動きを示すことが確認できた．今後は実験結果を利用し，評価式の定義・検証を行う．

参考文献

1) S. S. Wilks, “Mathematical Statistics”, Wiley International Edition, 1962

2)市野学，矢口博之，野中武志，“ 幾何学的厚みに基づく相関係数 ”，電子情報通信学会論文誌，Vol.J85-A，No.4，pp.490-494，2002

3) Y. Ono, M. Ichino, “A New Feature Selection Method to Extract Functional Structures from Multidimensional Symbolic Data, IEICE Trans. Inf. & Syst.,vol.E81-D, NO.6 June, 1998