よって地域人口等との関係が付けられる．したがって，各種の統計の結果が領域，或いは位置に対応するコードを持つならば，それを共通コードとして結合が可能になる．

今後，コンピュータがさらに大容量化し高速化したとしても，増大する一方のデータを全て結合したデータベースを作ることは実用的ではない．個々のデータのメンテナンスをも考えると，大型のデータベースはおそらく実用にはならないであろう．むしろ，メンテナンスは各統計の担当機関にまかせ，共通コードを利用して必要な属性のみ引出して使うという考え方が現実的であろう．

この様な観点から，地理情報を共通コードとしたときの統計解析の方法を検討している．

多次元クラスタリングとそれに伴う多次元異常値の摘出

川 合 伸 幸

1. はじめに

多次元データを解析する場合，まずクラスタリングすることによって，めざましい知見を得ることがある（河口（1977））．またそれに伴い，どの個体に異常値（outlier）の可能性があるかを知ることも重要である．ここで述べる手法は探索的データ解析の道具として位置づけられるが，多次元での異常値の探索を扱った文献としては，Gnanadesikan（1977）がほとんど唯一のものである．多次元異常値の摘出というと，機械的に取り出して捨ててしまうような印象を与えるが，異常値についてはその異常の理由が追求され，またなぜ他の個体と違っているかを考え，新しい知見に結びつける態度が重要と思われる．

2. 多次元異常値の摘出法

異常値の典型的な型として，

A：他の多くの個体と同一平面上にあってはずれているもの
B：他の多くの個体と同一平面上からはずれて存在しているもの

を考える（鷲尾・大橋（1989））．A, Bを区別するのに単なる主成分分析ではAはわかってもBはわからない．そこで，射影をすると必ず2次元平面上に再現された距離は短くなるという性質を利用し，短くなった程度を矢印の相対的な長さで示すことにする．これを“風向プロット”と呼ぶことにする．矢印は前記の性質によりすべて外側を向く．Aの型の異常値では絶対値の小さな風向プロットしか持たないのに対し，Bの型の異常値では四方八方に伸びた風向プロットを持つであろう．

3. データへの適用

2の方法を，去年も報告した書痙の分類データに適用してみた．しかし，これだけでは十分であるとは言えず，さらに多くの実践を通じて有効性を確認する必要があると思われる．

参 考 文 献

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York（丘本 正，磯貝恭史 訳（1979）．日科技連，東京）．
河口至商（1977）．『多変量解析入門II』，森北出版，東京．
鷲尾泰俊，大橋靖雄（1989）．『多次元データの解析』，岩波書店，東京．