

0.547を得ている(Tanemura (1979, 1991)). 今回いくつかの r に対する球帽のランダム逐次充填密度を計算機シミュレーションによって求め、 D_r の r 依存性を調べた。その際、 r として単位球面における最密規則配置のいくつかの n ($n=2, \dots, 32$)に対応する既知の値を選んだ。これによって規則配置に対する充填率も同時に得られる。その結果、比較的少数個の球帽の充填においてすでに極限密度 D_0 に近い値が得られることが示された。また、この問題では隙間の無い完全充填が達せられていることが必要で、そのために球面のVoronoi分割が有用であり、今回の結果はすべてそれに基づいている。

次に、海底に生息する放散虫の骨格構造を説明するモデルを考察した。放散虫の骨格は多角形による球面分割の構造を示しており、各多角形の面積がほぼ等しく大多数が六角形となる特徴を持っている。この構造では、球面ネットワークの稜の全長について最小性を満たすとの仮説を設定することができる。そのような構造は上述のランダム逐次充填の球帽系に対するVoronoi分割ネットワークでは得られないことは明白であった。そこで、球面上の点系のある初期配置から出発して徐々に位置を調節することによって上記の配置を実現する新しい幾何学モデルを考案し、そのアルゴリズムを提出した。

報告では、二つの放散虫データに対して上記のモデルのシミュレーション結果がそれぞれ極めてよく一致することを述べた。また、われわれのモデルが一般的に任意の n に対する最適点配置を得るのに有効であることも指摘した。

参 考 文 献

- Tanemura, M. (1979). On random complete packing by discs, *Ann. Inst. Statist. Math.*, **31**, 351-365.
Tanemura, M. (1991). On the density of strict random complete packing by discs, *Ann. Inst. Statist. Math.* (submitted).
Tóth, L.F. (1972). *Lagerungen in der Ebene, auf der Kugel und im Raum*, 第5, 6章, Springer, Heidelberg.

統計解析における地理情報の利用

馬 場 康 維

コンピュータの大容量化と処理速度の高速化に伴い、各種のデータが蓄積され、多くの変数の組合せや大量のデータによる統計分析が可能になってきたが、実際に種々の統計によるデータを組合せて使おうとすると、データの結合が難しいという問題がしばしば起きる。

社会現象を扱う統計解析では異なったソースのデータを結合して用いることがよく行なわれる。たとえば、なんらかの指標としてある地域の人口当りの生徒数を用いたいときには、文部省の生徒数の統計と総務庁の人口統計を用いる必要があるが、結合して用いるのは簡単ではない。こういった場合、地理情報がデータ結合の有効な情報として活用できる。

通常の統計情報のみならず様々な情報の蓄積がなされている。数値地理情報もその中の一つである。近年の画像処理技術の普及には目覚ましいものがあるが、官庁、民間を問わず、地図の利用が盛んになされており、各種の数値地理情報が集積されている。たとえば、地形等の自然的条件、行政区画、道路、鉄道、公共施設等を数値情報として持つ大規模なデータとして国土数値情報があり、河川や行政界等の位置座標、公共施設や文化財の所在を表す位置座標、地図をメッシュに分割したときのメッシュコード等のデータが蓄えられている。

地理情報の利用形態として一般的なものは、その情報をもとにして地図を描くことや各種の地域属性を地図上に表現することであろう。しかし、地理情報はもう一つの別の観点からの利用も可能である。

たとえば、国土数値情報の3次メッシュにより日本の国土は約1 km四方の小区画に分割される。ところで、地域の人口、学校数等、ある領域に属するデータはメッシュコードをIDとして持つ領域の属性と考えることができる。また、学校等位置座標を持つものも、それが存在するメッシュを指定することに

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抽出

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。多次元異常値の抽出という、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. 多次元異常値の抽出法

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、去年も報告した書齋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.