隣接関係に基づくクラスター分析 —ドロネー図を利用した隣接関係の記述—

氏名：大内勇輝

指導教員：大沼進

クラスター分析は、多次元空間中のヒトやモノなどの個体を、その類似度に基づいてクラスターとよばれるカテゴリーに分類する多変量解析手法である。その応用範囲は、天文学、コンピュータ科学、生物学、社会科学、心理学など多岐にわたり、多変量データに潜在する個体の集合の性質を探索的に調べる方法として広く利用されている。

クラスター分析はこれまでに多くの手法が提案されているが、その多くは個体間の類似度や非類似度に基づいて分類を行う。たとえば階層的クラスター分析に属する最短距離法、最遠距離法や Ward 法、非階層的クラスター分析に属する k-means 法や Fuzzy C-means 法などがよく知られている。

しかしながら、近年画像処理や機械学習の分野でよく使われるようになったスペクトラルクラスタリングは、従来の方法とは大きく異なる部分がある。スペクトラルクラスタリングはグラフ理論に基づいて構築され、個体間の隣接関係を記述する隣接行列を利用して分類を行う。隣接行列にはさまざまな定義があり、 k 近傍や ϵ 近傍がよく用いられてきた。

本研究では、 k 近傍や ϵ 近傍よりも適応的に隣接関係を記述できると考えられるドロネー図に基づく隣接行列の利用を提案した。ドロネー図は空間を分割する図形の一つで、個体を表す点からの距離が他の個体からよりも近い点の集合であるボロノイ図の双対図形として定義される。このためドロネー図は、隣接関係に基づくスペクトラルクラスタリングとの親和性が高く、隣接行列の基盤として有効であると考えられる。本研究では、 k 近傍、 ϵ 近傍、そしてドロネー図（重みなし）、ドロネー図（正規重み）により定義される隣接行列を用いて、人工データにスペクトラルクラスタリングを適用するシミュレーションを行い、ドロネー図に基づく隣接行列の有用性を検証した。