氏　　　　名	高井 勉
学位(専攻分野)	博士(統計科学)
学 位 記 番 号	総研大甲第 1995 号
学位授与の日付	平成 30 年 3 月23 日
学位授与の要件	複合科学研究科 統計科学専攻 学位規則第6条第1項該当
学 位 論 文 題 目	空間点パターンのグラフィカルな分類方法に関する研究
論 文 審 査 委 員	主　査　教授 中野 純司 教授 田村 義保 准教授 加藤 昇吾 理事長 椿 広計 独立行政法人統計センター 准教授 元山 斉 青山学院大学 経済学部

(別紙様式 2)
(Separate Form 2)

論文の要旨

本論文では、空間点パターンをグラフィカルに分類する方法として AG-curve（AGglomerative-curve）を提案した。これは凝集型階層的クラスター分析（Agglomerative Hierarchical Clustering Algorithm）を応用した方法である（以降は AHC 手法と呼ぶ)。

本論文は、

・AG-curve を用いた空間点パターンの分類方法の提案

・空間点パターンの分類に適した AHC 手法の探索

・既存のグラフィカルな分類方法に対する AG-curve の優位性の検証

・AG-curve の数理的性質の明確化

を目的に書かれた。

自然科学分野での空間点パターンは 2 次元の事例が多い。しかし多変量（多次元）を取り扱う分野での利用を考え、先行研究で提案された手法から本論文で提案する手法まで、各種統計量の導出に当たっては多次元を基本とした。

まず空間点パターンを研究対象とする分野、および様々な空間点パターンを紹介した。本論文では、空間点パターンを、ランダム型・集中型（凝集型)・規則型の 3 種類に分類した。その中のランダム型を厳密に定義するために用いる完全空間ランダム性（CSR, Complete Spatial Randomness）は、分類の基準となるもので最も重要である。これは時系列解析や信号処理における、ホワイト・ノイズ性に相当するものである。また、既存の手法と提案手法を統一的に比較するために、共通の分類対象として用いる 3 種類の空間点パターン（2 次元）を示した。これらは全て現実の空間点パターンである。

空間点パターンの分類方法の中には近接グラフと関係の深いものが多い。既存のグラフィカルな方法および提案手法である AG-curve の基礎をなす、近接グラフの辺長の分布を先行研究に従い導出した。近接グラフとは、点の近さを表す「近接性」によって構築されるグラフである。本論文で取り上げた近接グラフは、第 1 近隣木、第 k 近隣木、相互最近隣対（Reciprocal nearest neighbor pair)、MST（Minimum Spanning Tree）である。領域内に強度 λ で一様ランダムに配置された点過程に対する辺長の分布を対象とした。多次元で導出したのち、具体例として 1 次元・2 次元・3 次元の結果（確率密度関数、期待値、標準偏差）を示した。

本論文では、提案手法である AG-curve と比較するため、空間点パターンの探索的データ解析に用いられる、既存のグラフィカルな分類方法の使用方法と特徴を示した。これらは近接グラフ(Proximity graph)に基づくものと､2 次特性(Second-order characteristics)に基づくものに分けられる。前者として G 関数、F 関数、J 関数を、後者として K 関数、L 関数、pair correlation 関数を取り上げた。空間点過程の 2 次特性とは、2 点の相互作用、あるいは空間上の関係が空間内でどのように変化するのかを記述する量である。各分類手法の説明の後に、その分類手法を用いた具体例を示した。分類対象の空間点パターンが、ランダム型・規則型・集中（凝集）型のどのタイプに属するのかをグラフィカルに判定する内容である。ここでは分類対象として 3 種類の現実の点パターンを用いた。

提案手法である AG-curve は AHC 手法の併合距離とクラスター数の間の関係を示して

いる。そこで AHC 手法の概略を Lance and Williams（1967）の組合せ的な手法を用いて示した。次に多くの AHC 手法が持つ単調性に焦点をあてた。単調性とはクラスター化の進展とともに併合距離が単調に増加することを意味する。この単調性は AG-curve からの要請である。単調性の説明のため、様々な距離がもつ性質を示したのち、単調性の判別条件の基礎となる強メトリック不等式を示した。最後に、単調性が保証されている AHC 手法として、具体的に最短距離法・最長距離法・（重み無し）群平均法を取り上げ、それぞれの併合距離の定義とクラスター化の特徴を示した。さらに比較のため、単調性が保証されていないセントロイド法も取り上げた。

本論文の主要な部分として、提案する AG-curve の定義と作図手順を示したのち、空間点パターンの分類に適した AHC 手法を探索した。AG-curve には単調性が期待されているので、その基礎となる AHC 手法にも単調性が期待されている。したがって探索する AHC 手法の候補として、単調性が保証されている、最短距離法・最長距離法・群平均法の 3 手法を選択した。結果として最短距離法が選択された。最短距離法（single-linkage）を用いた AG-curve を AGsi-curve と呼ぶ。本論文の残りの部分では、AGsi-curve に的を絞って議論を進めた。

次に AGsi-curve の分類特性を、既存の手法である G 関数および K 関数と比較した。それらに対する優位性を示すために計算機実験を活用した。最後に、AGsi-curve は MST の辺長の経験分布関数と等価であることを示した。補足として MST の辺長の経験分布関数を 関数とし、この 関数を用いた空間点パターンの分類事例を示した。

明らかにされた「AGsi-curve と MST の辺長との関係」を深く研究するため、CSR 点過程に対する AGsi-curve の振舞いを MST の辺長の順序統計量を用いて説明することを試みた。MST の辺長が従う確率分布は解析的には求められないが、幾つかの近似分布が提案されている。その中から本論文では Roberts（1968）の近似と Watanabe（2008）の近似を取り上げた。Roberts（1968）の近似分布の順序統計量を用いた場合を AGsi-curve の近似分布 U とし、Watanabe（2008）の近似分布の順序統計量を用いた場合を AGsi-curve の近似分布 L とする。これらの近似分布を多次元で導出した。AGsi-curve の近似分布 U の導出においては MST の辺長に関する Roberts（1968）の近似分布の順序統計量を用いた。また、AGsi-curve の近似分布 L の導出のためには MST の辺長に関する Watanabe（2008）の近似分布の順序統計量を用いた。

導出した近似分布 U と近似分布 L の特性を明らかにするため、近似分布の期待値の強度依存性に着目し、モンテカルロ法により推定した厳密分布の期待値と比較した。次に AGsi-curve の近似分布の導出に用いた、MST の辺長の近似分布である Roberts（1968）と Watanabe（2008）についても、期待値の強度依存性に着目し、モンテカルロ法により推定した厳密分布の期待値と比較した。最後に、明らかにされた AGsi-curve の近似分布の特性と MST の辺長の近似分布の特性を比較し考察した。AGsi-curve の近似分布は、その基になった MST の辺長の近似分布と深い関係があることが示された。また、厳密分布の期待値と近似分布の期待値の相対的な位置関係は、MST の辺長の境界効果で説明できることが分かった。

(別紙様式 3)
(Separate Form 3)

博士論文審査結果の要旨

本学位論文の審査は、2018 年 1 月 18 日に５名の審査員によって行われた。

空間点パターンをランダム型・規則型・凝集型の 3 種類にグラフィカルに分類する方法として、これまで G 関数、K 関数をはじめとする様々な方法が提案されてきた。本論文では新たな分類方法として、凝集型階層的クラスター分析法を応用した AG-curve を提案し、その１種である AGsi-curve に焦点を当て､その分類性能と数理的性質について論じている。本論文の特色として、既存の方法および提案する方法の分類性能に関して、その基礎を成す近接グラフおよび階層的クラスター分析法に遡って説明していることがあげられる。これにより各方法の長所・短所が理解し易くなっている。

本論文では、現実の点パターンおよび人工的な点パターンを用いて、既存の分類方法に対する AGsi-curve の優位性が示されている。具体的には、既存の方法では分類できない 2 種類の点パターンを分類できることや、境界効果に対するロバスト性などである。また、AGsi-curve の数理的性質は Minimum Spanning Tree の辺長の順序統計量を用いて説明できることから、AGsi-curve の 2 種類の近似分布を新たに導出している。モンテカルロ法により推定した厳密分布と比較することにより、2 種類の近似分布は限定された強度の範囲内では、近似モデルとして有用であることを示した。

一方、AGsi-curve は既存の分類方法と比較すると境界効果に対してロバストであることを示したが､モンテカルロ法により推定した AGsi-curve の厳密分布と導出した近似分布を比較した結果､強度が小さくなると AGsi-curve の境界効果も無視できないことが明らかとなった。AGsi-curve の境界効果は、その基礎を成す Minimum Spanning Tree の辺長の境界効果によるものであることを示したのち、先行研究を発展させ、簡易的なモデル化と計算機実験を用いて、Minimum Spanning Tree の辺長の境界効果を定量的に評価した。

本論文の貢献は、凝集型階層的クラスター分析法に基づいて空間点パターンをグラフィカルに分類する方法として AGsi-curve を提案し､その数理的性質を明らかにしたことである。既存の G 関数を用いた場合では区別できない点パターンの区別ができることやλ（強度）が大きい場合に K 関数より規則型とランダム型の分類が明瞭にできること、境界効果に対して頑健であるなどの優位性があることなどは重要な結果である。分類だけでなく、ランダム性の検定のために用いるように拡張できる可能性がある等、空間点パターンに関する諸問題を解決するために用いることができるものと考える。

公開論文発表会において、出願者は研究内容を分かり易く説明し､「最短距離法を選択した理由」「境界効果と境界補正」などの質問に対する回答も適切であった。本論文の結果は統計科学的に有用であると考えられ、学位論文としてふさわしいものとして、審査委員全員一致で合格と判定した。