

学位論文の要旨

Abstract of Thesis

研究科 School	環境生命科学研究科 Graduate School of Environmental and Life Science
専攻 Division	環境科学専攻 Division of Environmental Science
学生番号 Student No.	77430302
氏名 Name	北西 由武 KITANISHI Yoshitake

学位論文題目 Title of Thesis（学位論文題目が英語の場合は和訳を付記）

医薬品関連データの俯瞰と仮説立案を目的とした位相的データ解析の活用とその有用性の考察

学位論文の要旨 Abstract of Thesis

医薬品業界においてもビッグデータ活用は急速に浸透してきており、ビッグデータ解析の方法論や解析プロセスの開発も進み、多種多様な解析方法が存在する。中でも、解析を実施する初手として、分類は効果的な解析手法であると考え、分類の目的はデータを俯瞰し、データの特徴を把握するためであり、クラスタリング手法がよく用いられてきた。しかしながら、従来のクラスタリング手法（例えば階層的クラスター分析）では、それらデータの特徴を空間的に捉えること、また、データの更新と増加に対してロバストかつ安定的に可視化すること、さらにはノイズの影響を取り除いて真の構造を可視化することは難しい課題である。そこで新しい可視化の技術として位相的データ解析マッパー（Topological Data Analysis Mapper：TDA Mapper）が注目されている。本稿では、これら医薬品に関するビッグデータを可視化する方法としての TDA Mapper の適用場面を考察し、多様なデータを組み合わせることによるデータ活用の提案、さらには、その関連技術を組み合わせることで新奇的な解析、可視化方法について、3つのアプローチを提案した。

- ① TDA Mapper をアソシエーション分析から得られる各指標に適用することで、視覚的に結果の関連性を把握することを提案した。アソシエーション分析から出力される結果としての3つの指標である支持度、確信度、及びリフト値といった多次元の指標を、効果的に可視化する方法として TDA Mapper を用いた。具体的にはレセプトデータを基に、Patient Journey を表現する方法を提案した。多次元指標を総合的に判断し、連続的な形状

として捉え、その特徴から適切な解釈をすることを可能にし、結果的に TDA Mapper の適用は非常に効果的であった。

- ② 解析対象として従来から扱われてきた構造化された定量的データ（本論文では、化学記述子データ）に加えて、非構造化の定性的データといった新たなモダリティデータ（本論文では、医薬品に関するテキストデータ）も解析対象とし、医薬品情報データに TDA Mapper を適用した。結果的に定量的データと定性的データの統合データについて、適応疾患に基づき連続的な形状として捉えることができおり、分類結果から様々な解釈を行うことができた。
- ③ クラスタリング手法の課題である距離（非類似性）に着目し、特に超多変数の場合の距離の算出方法について提案した。特に、クラスタリングの距離を算出する際には、変数が多いほど、さらにノイズが多いほど、距離行列が攪乱されてしまう。すなわち、ビッグデータに対して、従来のクラスタリング方法（距離算出の方法）により、得られた結果は解釈できない場合もあった。そこで、変数情報をサブセットで捉えて距離行列を算出する COSA（Clustering Objects on Subsets of Attributes）アルゴリズムで算出した距離行列をもとに、可視化手法として TDA Mapper を組み合わせて用いることを提案した。結果的に、ノイズを除去し、攪乱されたデータを整理し、真の構造に近い状態にした上で、そのデータを形状として捉え、ロバストに可視化できた。

本論文を通じて、多次元指標の可視化、新たなモダリティデータへの対応、ノイズや多次元データへの対応を通じて、TDA Mapper の有用かつ効果的な利用活用方法を提案した。これらの課題解決方法は、本論文の適用事例として示したとおり、医薬品の研究開発のいくつかの局面において、大いに役に立つ可能性がある。ビッグデータ活用においては、データを一つ一つ確認していくことは現実的ではなく、データ全体を捉え、効率的にデータを俯瞰し、その上でポイントを絞り、データを細かく確認することで効果的な仮説立案を可能にする。すなわち、医薬品ビッグデータを解析する上で、本論文で提案した TDA Mapper のビッグデータへの適用とその関連手法の組み合わせは、非常に有効な手段であることがわかった。