

氏名	北西 由武
授与した学位	博士
専攻分野の名称	理学
学位授与番号	博甲第 6646 号
学位授与の日付	2022年 3月 25日
学位授与の要件	環境生命科学研究科 環境科学専攻 (学位規則第4条第1項該当)
学位論文の題目	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
論文審査委員	教授 栗原考次 教授 坂本亘 教授 飯塚誠也 准教授 石岡文生
学位論文内容の要旨	
<p>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、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。その中でも、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られる。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る。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は難しい問題である。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。本論文では、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の活用、さらには、その関連技術を組み合わせることによる新奇的な解析、可視化方法について、3つのアプローチを提案した。まず、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。すなわち、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た TDA Mapper の適用は非常に効果的であった。次に、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に加え、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、テキストデータ）も解析対象とし、医薬品情報に 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。最後に、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、超多変数の場合の距離の算出方法について提案した。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、さらにはノイズが多いほど、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ものだったのかもしれない。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 を組み合わせることを提案した。本論文では、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医薬品の研究開発のいくつかの場面において、大いに役に立つ可能性がある。</p>	

論文審査結果の要旨

本論文では、医薬品に関するビッグデータの分析に対して、位相的データ解析の一つの方法である TDA (Topological Data Analysis) Mapper を活用し、多様なデータを組み合わせることによる 3つのアプローチを提案し、その有効性を検証している。

最初のアプローチは、アソシエーション分析から得られる 3つの指標に TDA Mapper を適用し、多次元指標を総合的に連続的な形状として捉えることによりその特徴から適切に解釈する方法を提案している。具体的には、医療機関が患者に対して行った医療行為や傷病名などの詳細を示した 560 万人のレセプトデータに対して、アソシエーション分析により得られる支持度、確信度、リフト値の 3つの指標に TDA Mapper を適用することにより、うつ病に関連する病歴や症状などのペイシェント・ジャーニー (Patient Journey) を表現することを可能にしている。

次のアプローチは、医薬品情報がテキストデータとして記述されている Wikipedia 及び医薬品情報が化学記述子データの化合物情報として格納されている DrugBank の 2種類のデータを定量化し、TDA Mapper を適用する方法を提案している。すなわち、テキストデータ及び化合物情報の化学的な構造情報に基づき、医薬品の位相的な構造を求める方法を提案している。提案方法により、適応疾患に関連する連続的な形状として捉えることができ、分類結果から医薬品の様々な解釈及びその分類結果に基づくドラッグリポジショニングが可能となる。

最後のアプローチは、ビッグデータに対してクラスタリングを行う際の課題である距離 (非類似性) に着目し、超多変数の場合やノイズが多い時の距離の算出方法について提案している。すなわち、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、可視化のための TDA Mapper を組み合わせて用いる新しい方法を提案するとともに提案方法が有効であることを示している。

これらの研究成果は、2編の査読付き参考論文、3編の査読付き関連論文、15件の口頭発表などにおいて公表している。これらの研究は、製薬業界におけるデータサイエンスに関する新分野としての理論と応用に大きく貢献すると考えられる。以上により、本論文は博士 (理学) に値すると判断した。