

Title	大規模動的データに対する探索的分析法に関する研究
Author(s)	福井, 健一
Citation	大阪大学, 2010, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/57621
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉 大阪大学の博士論文について 〈/a〉 をご参照ください。

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	福井 健一
博士の専攻分野の名称	博士(情報科学)
学位記番号	第 24093 号
学位授与年月日	平成22年3月23日
学位授与の要件	学位規則第4条第2項該当
学位論文名	大規模動的データに対する探索的分析法に関する研究
論文審査委員	(主査) 教授 沼尾 正行 (副査) 教授 森田 浩 教授 石井 昭博 教授 谷田 純 教授 八木 厚志

論文内容の要旨

長期に渡る大規模データの全貌が視覚的に把握できれば、大局的な現象の直感的な理解や、人がデータを探索的に分析するための基盤となる。本研究ではデータを「よく観察する」行為を支援する目的で、可視化を中心としたダイナミズムの探索的分析法に関する研究を行った。

まずデータ全体の特徴を表すひとつの基本的な要素として、類似データからなるクラスタが挙げられる。クラスタ数の変化への対応や、時間変化に対するクラスタの一貫性の問題に対して、本研究では自己組織化マップ(SOM)のトポロジーにより特徴空間上のデータの分布を捉える特長を活かし、トポロジーの一方に順序関数を導入することで、時間方向の分布の変化を捉える時系列SOMを提案した。人工データを用いてクラスタの変化(クラスタの移動、拡大・縮小、融合・分離)が提案法により可視化されることを確認し、新聞記事から抽出したトピックの推移や燃料電池の損傷過程がクラスタマップから読み取れることを示した。

次に、クラスタ分析が類似データの集合に着目しているのに対して、オブジェクト間の関係と全体との関係を分析する方法にネットワーク分析がある。本研究では、明示的な二項関係を持たない対象に対してネットワーク分析を適用した。具体的には、燃料電池の損傷計測Acoustic Emission(AE)信号データの類似近傍から構成された近傍ネットワークにも複雑ネットワークの特徴があり、そのネットワーク構造の変化を調べることで損傷のダイナミズムの一側面を理解できることを示した。

また探索的な分析では、通常、抽出された結果は利用者が解釈することになる。例えば、抽出されたクラスタの意味は、後付けで利用者が解釈するといった「解釈問題」が存在する。本研究ではこの解釈問題へのひとつのアプローチとして、トピックの解釈を例に挙げ、人手により抽出したトピックを用いて機械抽出したトピックの解釈可能度を定量化する方法を提案した。新聞記事から抽出したトピックに適用し、従来の定性的な知見と一致することを確認した。

上述のクラスタマップや近傍ネットワークの精度、つまり正確に類似データ同士が近傍に配置されているかどうかは、利用者の探索的な分析に影響を与える。そこで本研究では、損傷によるAE波に対して、周波数スペクトル全体の分布の近さを類似度とすることでクラスタマップの分類・可視化の精度向上を図った。模擬AEデータを用い

て分類精度が向上していることを確認するとともに、燃料電池の損傷過程がより詳細に把握できることを示した。

論文審査の結果の要旨

今後、Webのみならず実世界の情報を含んだセンサネットワークやライフログなどの長期に渡る多次元データが得られるようになり、かつ未知の対象を扱う状況に直面する。本論文は、まずは大局的な現象の直感的理解と、利用者がデータを探索的に分析するための基盤として、可視化を中心としたダイナミズムの探索的分析法に関して、既存手法を拡張するとともに、未知の応用分野への適用を行い有用な知見を得ている。

まず、データ全体の特徴を表す基本要素である類似データからなるクラスタに関して、本論文では自己組織化マップ (SOM) のトポロジーにより特徴空間上のデータ分布を捉える特長を活かし、トポロジーの一方に順序閾数を導入することで、時間方向の分布の変化を捉える時系列SOMを提案している。提案法はクラスタ数の変化、時間変化に対するクラスタの一貫性に関して従来法よりも優れている。人工データを用いてクラスタの変化（クラスタの移動、拡大・縮小、融合・分離）が提案法により可視化されることを確認するとともに、新聞記事から抽出したトピックの推移や、燃料電池の損傷過程がクラスタマップから読み取れることから、提案手法が実問題にも十分適用可能であることを示している。

次に、オブジェクト間の関係と全体との関係を分析するネットワーク分析に関して、本論文では明示的な二項関係を持たない対象に対してネットワーク分析を適用している。具体的には、燃料電池の損傷計測Acoustic Emission (AE) 信号データの類似近傍から構成された近傍ネットワークにも複雑ネットワークの特徴があり、そのネットワーク構造の変化を調べることで損傷のダイナミズムの側面を理解できることを示している。

一方、探索的な分析において抽出された結果の意味を利用者が解釈する解釈問題に対して、本論文ではトピックの解釈を例に挙げ、人手により抽出したトピックを用いて機械抽出したトピックの解釈可能度を定量化する方法を提案している。従来、機械抽出トピックの評価は人手による後付け解釈もしくは、人手による抽出トピックとの一対一対応によって行われていたが、本評価法は多対一での対応を考慮した評価が可能となっている。そして、新聞記事から機械抽出したトピックを提案法により評価し、従来の定性的な知見と一致することを確認している。

また、クラスタマップや近傍ネットワークの精度、つまり正確に類似データ同士が近傍に配置されているかどうかは、利用者の探索的な分析に影響を与える。本論文では、損傷によるAE波に対して、周波数スペクトル全体の分布の近さを類似度に導入することでクラスタマップの分類・可視化の精度向上を図っている。模擬AEデータを用いて分類精度が向上していることを確認するとともに、燃料電池の損傷過程がより詳細に把握できることを示している。

以上のように、本論文は情報科学の発展に寄与する研究である。よって、博士（情報科学）の学位論文として価値のあるものと認める。