

クラスタリングに基づく情報の検索と視覚化

堀田, 政二

<https://doi.org/10.15017/1398256>

出版情報 : 九州芸術工科大学, 2001, 博士 (工学), 課程博士
バージョン :
権利関係 :

第9章

結論

本論文では、クラスタリングを利用した情報の検索と視覚化について研究した結果をまとめた。本研究によって得られた研究成果の概要は以下のとおりである。

第1章では、本研究の背景と扱っている問題を示し、あわせて論文の概要について述べた。

第2章では、類似度行列からファジークラスタを逐次に抽出する方法を提案した。ファジークラスタリングを最適化問題として定式化し、ファジークラスタ抽出問題が固有値問題に帰着されることを示した。これはグラフスペクトル法と総称される方法の一つである。各データの残存率を類似度行列の要素に掛けていくことにより、クラスタを逐次に抽出する方法を示した。抽出処理はクラスタの凝集度の変化に基づいて、ある程度の大きさのクラスタが抽出されたら終了するようにした。2次元データを使った簡単な実験によりノイズデータに頑健なクラスタが抽出されることを確認した。実データへの応用としてグレイスケール画像のセグメンテーションとビデオのセグメンテーションを行い、良好な結果を得た。

第3章では、隣接行列に基づいて点集合をファジークラスタ(ファジー部分グラフ)に分割するグラフスペクトル法の1種を提案した。まず最初に、第2章で提案したファジークラスタ逐次抽出法を点と辺に重みを持つ無向グラフに適用し、続いてそれを有向グラフと2部無向グラフに拡張してグラフ構造データからファジークラスタを逐次に抽出する方法を提案した。 k 平均法との比較実験を行い、提案手法がノイズに対しロバストであることを示した。提案したクラスタリング法は、点の接続関係だけに基づくものであり、通常のように点自体の付随的特性(物理特徴量など)だけに基づく方法とは対極的である。

第4章では、第3章で提案したファジークラスタ逐次抽出法を、複数のグラフが合成した複雑なグラフに適用できるように拡張した。複雑なグラフ構造データでは

固有値問題に帰着できないため、べき乗法を一般化した反復法による抽出法を導いた。はじめに、無向グラフと2部無向グラフ、および有向グラフから反復法に基づいてファジークラスタを逐次に抽出する方法を示し、有向グラフと2部無向グラフが合わさった複雑なグラフからファジークラスタを抽出する方法を示した。また、これらの反復法が収束することを証明した。例題として簡単な混成グラフを使った実験を行い、提案手法の有効性を確認した。提案手法は複雑なグラフ構造データへの拡張が容易であるので、ビデオのような、色、音声、テキストなどを統合したデータへ応用するのが今後の課題である。

第5章では、データ行列に基づき、数量化3類と数量化4類によってデータ構造を視覚化する方法を示した。また、数量化4類を重み付きデータに適用できるように拡張した。これらの手法は、最適解が解析的に求まる固有値問題に帰着され、一種のグラフスペクトル法であることを示した。数量化3類と数量化4類についてデータ行列に基づく視覚化に関する比較実験を行い、両者の長所と短所を明らかにした。データ数が少ない場合の実験例では、数量化3類よりも数量化4類の方が互いの距離関係に即したデータ配置を得られることを確認した。一方、データ数が多い場合、クラスタリングによって得られたメンバシップ値を利用して数量化3類でデータを配置すれば、数量化4類よりも良好な配置が得られ、計算時間も少なくすむことを確認した。さらに、クラスタリングの結果を重み付き数量化4類により逐次展開して表示する方法を提案し実験例を示した。

第6章では、グラフ構造データをファジークラスタリングし、得られたメンバシップに基づいてハンティング検索する方法とブラウジング検索を行う方法を提案した。ファジークラスタリングの方法としては、データ構造が無向グラフや有向グラフ、あるいは2部無向グラフである場合には、グラフスペクトル法の一つである第3章の方法を、混成グラフの場合は反復法である第4章の方法を用いた。提案手法のハンティング検索について実験を行い、提案手法がLSI法と同様にクエリとの表面的な一貫性でなく潜在的な関連性を反映した検索が行えることを確認した。また、グラフ構造データの表示法としては、第5章の数量化3類による視覚化法をグラフデータに応用してデータのクラスタ構造を視覚化し、ブラウジング検索や視覚的なリンクの推薦などに利用した。本論文では比較的少数のデータの検索や表示を扱ったが、データ数が大量になると階層化などの構造化を行わなければ検索も表示も効率が落ちると思われる。ここで提案した方法を階層的クラスタリングに拡張して大量データに応用するのが今後の課題である。

第7章では、画像検索を高速化するためのフィルタリング法として、特徴ベクト

ルの次元削減に基づく距離の不等式とクラスタリングに基づく三角不等式とを利用する手法を提案し、実験により画像検索の時間が短縮できることを示した。提案手法における次元削減とクラスタリングを階層的に拡張することが今後の課題である。

第8章では、画像の近似 k NN検索法として、2次形式の標準化と主成分分析による次元圧縮、およびクラスタリングに基づく方法を提案し、500枚の画像データで実験した結果、元の次元での全探索とほぼ同じ結果を出力するのに約10倍、第7章で提案したフィルタリング法と比べても約3倍高速化することができた。また、次元削減によるフィルタリングを組み合わせることにより更に約2倍高速化できることを実験で示した。