

科学研究費補助金研究成果報告書

平成 22 年 5 月 31 日現在

研究種目：若手研究(B)

研究期間：2008 ～ 2009

課題番号：20700091

研究課題名（和文） 大規模な階層構造データに対する探検型可視化手法の開発

研究課題名（英文） Graph-Based Visualization Method for Large Hierarchies

研究代表者

井上 悦子 (INOUE ETSUKO)

和歌山大学・システム工学部・助教

研究者番号：50452537

研究成果の概要（和文）：

本研究では、先行研究で開発した大規模なクラスタリング結果の任意の部分領域をグラフを用いてインタラクティブに可視化する探検型可視化手法を発展させ、より実用的な可視化システムとして完成度を高めた。具体的には、実用的な応用分野の一つとしてニュース記事等の文書のクラスタリング結果を対象として、クラスタリング結果ビューアを構築した。また、本システムに複数の Web ブラウザからアクセスすることを目指として、サーバクライアント方式への拡張を行い、基礎的な仕組みを実現することで、サーバクライアント方式の実用可能性を確認した。

研究成果の概要（英文）：

In this research, we improved our previous result of “exploration visualization method,” which visualizes a portion of large scale clustering results interactively like “exploring” the data, to be more practical visualization methodology. Specifically, we first developed a system to visualize a clustering result of news articles to overview the whole news variation as a good example of the applications. Next we improved our system to be server-client type system to endure many accesses from web browsers. We propose the basic mechanisms for that, and through implementing them we confirmed partly the possibility to be realized as a web services in the Internet.

交付決定額

（金額単位：円）

	直接経費	間接経費	合 計
2008 年度	500,000	150,000	650,000
2009 年度	600,000	180,000	780,000
年度			
年度			
年度			
総 計	1,100,000	330,000	1,430,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：可視化

1. 研究開始当初の背景

近年、インターネットの広がりによりコンピュータで扱うことのできる情報量が格段に増加しており、今後は大規模データを処理する技術が重要になることは疑いない。大規模データを対象とした分析技術の発展も重要課題である一方で、大規模データの分析結果などを十分に理解するためには、分析結果を分かりやすく可視化することが非常に重要である。

本研究では、対象データの要素間に類似度などで測られる数値化された関係が定義される場合を想定し、これらの要素間の関係をクラスタリング手法を用いて分類した結果を可視化する問題を扱う。我々の先行研究では、全ての要素間に類似度が定義され、これが階層型クラスタリング法により処理された結果を、インタラクティブなグラフ表示機能を用いて、あたかもクラスタリング結果空間を探索しているかのような感覚で閲覧することができる「探索型可視化手法」を提案してきた。

しかしながら、この探索型可視化手法を実用化するにあたっては、システムの実行にあたって多くの計算機リソースを必要とするため使用できる場面が限定されること、一般の階層型クラスタリング結果のデータのみを対象とするため適用できるデータの範囲が限定されること、などの問題点が挙げられていた。

2. 研究の目的

本研究課題では、先行研究で試作した探索型可視化システムを実用化するための改良を行う。先行研究で試作したシステムは、階層構造を持つ大規模なデータに対して、任意の部分領域をグラフの形式で閲覧しながら、ユーザがインタラクティブに対象領域を移動させることにより、データ空間内を自由に探索するような新しい可視化手法を実現している。

研究開始当初の目的は、探索型可視化手法を適用できるデータ範囲が限られていた問題に着目したものであり、本手法の対象として、要素間の関係がより制限されたグラフ構造を持つデータ（グラフ構造を持つデータに対して、階層的にクラスタリングを行う手法も提案されている）を対象に含める、あるいは二分木として実現される階層型クラスタリング結果だけではなく、より一般的な木構

造が定義されている階層構造データを扱えるようにする、などの適用範囲の拡張を行うことを目標としていた。

しかし、本手法に関する検討を進めるうえで、まずは現在の可視化手法を何らかの形で実用化することで、本手法の有用性を検証することの方が重要であるという結論に至り、このため本研究課題では、実用例として(1)Web上に大量のデータが存在するテキストデータを扱ったクラスタリング結果の可視化システムを実現すること、および(2)Web上で分析結果を公開し、数多くのユーザがアクセスできるようにするためのサーバクライアント方式（サーバがある程度のリソースを持っていれば、クライアントの性能は低くても良いようにする）の実現を目標として研究を進めてきた。

3. 研究の方法

(1)本システムの実用例を示すために、先行研究で開発した試作システムの完成度を向上させることにより、ビューソフトウェアとして実用に耐えるよう拡充を行った。具体的には、試作システムの問題点であった、ユーザによるインタラクティブな操作に対応して変化した要素を表示グラフ内でみつけづらい点や、複数要素から成るクラスタ内の個々の要素の情報を把握することが難しい点など、いくつかのユーザインタフェース上の問題を解決することによりシステムの完成度を高めた。

そのうえで、本システムをWeb上に大量に存在するテキストデータのクラスタリング結果を閲覧するシステムへと特化させるよう実装を進めた。対象データとして、比較的クラスタリングによるデータの分類・階層化への要求が高いWeb上のニュース記事を選んだ。ニュース記事を形態素解析し、TF/IDF値によるキーワード抽出の後、コサイン尺度を用いて記事の階層型クラスタリングを行い、その結果を可視化するシステムを構築した（図1）。ニュース記事を可視化するためのインタフェースとして、各記事あるいは記事クラスタに含まれるキーワードを一覧表示する機能や、記事のジャンルにより色分けする機能などを充実させることにより、クラスタリング状況が把握しやすいインタフェース設計とした。

(2)システムのサーバクライアント化に関しては、Web上にある程度のリソースを持つサーバを設置し、ユーザが少リソース環境で

も快適にデータを閲覧できるような応用場
面を想定して、現在の仕組みを拡張する検討
を行った。クライアントを少リソースで実現
し、多くのクライアントによる並列利用を可
能とするためには、サーバとクライアントが
持つデータの切り分けが肝心である。クライ
アントには大量のデータを保持することが
できないので、必要最低限のデータ量のみを
保持する仕様にする必要がある。その一方
で、サーバに個々のクライアントの状態（閲
覧中のノード領域など）を保持してしまうと
接続クライアント数に限界を設けることにな
るため、できるだけ避けなければならない。
これらの制約を満たすような保持データの切
り分け方法、およびデータの通信プロトコル
を提案した。また、この提案手法を実現する
プロトタイプシステムを構築し、性能評価を
行った。

4. 研究成果

(1) 探検型可視化システムの実用例の構築に
関しては、ビューソフトウェアとしての完
成度を高めることができた。ユーザによる可
視化領域に対する操作（表示レベルを上げる、
下げる）に対して、グラフ表示の画面上で発
生するアニメーションがあっても、どの要素
が変化したのかなどの画面内の情報が理解
できるようになり、またユーザインタフェー
スの各所が改善された。これら改善を踏まえ
て、研究成果を電子情報通信学会論文誌に発
表した。

また、本システムをニュース記事のクラ
スタリング結果に特化させた可視化システム
を設計・実装した。また、本システムを Web
上の複数新聞社サイトのニュース記事デー
タを階層型クラスタリングした結果に適用
することで、システムの評価を行った。その
結果、大規模なクラスタリング結果に含ま
れている適切な部分や、不適切な部分を発見
することができることが検証され、ニュース
記事のクラスタ構成を把握することができる
ことが示された。本結果は、学生の卒業論文
としてとりまとめられた。

(2) システムのサーバクライアント化につ
いては、前述のとおり、サーバとクライ
アントに保持するデータの切り分け、および通
信プロトコルの確定を行い、大規模なデータを
サーバが保持している状態で少リソースな
クライアントにおいて探検型可視化が実現
可能になるような仕組みを提案することが
できた。この仕組みをプロトタイプシステム
として実装し、同一 LAN 内でサーバとクライ
アントを動作させて性能評価を行った。評価
の結果、クライアントが必要とするリソース
は十分に低いことが実証され、また、ユーザ



図 1. 開発したニュース記事クラスタ
リング結果可視化システム。グラフは
左上部に表示され、ニュースのジャン
ル毎に色分け表示される。下部にはク
ラスタリング結果の樹状図の全体像が
表示され、表示領域がわかる。右上部
は各記事や記事クラスタのキーワード
や記事タイトル等を一覧できるように
なっている。

による可視化領域への操作に対する反応速
度も十分に早いことが確認できた。一方で課
題点として、ユーザが短時間に多くの操作を
行った場合にはクライアントのグラフ表示
の遅延など、システムの反応が鈍くなること
が想定されるため、インターネットのような
遅延の発生しやすいネットワーク環境での
使用にあたっては、クライアントでのグラフ
表示に必要なデータを事前にキャッシュし
ておく機能が必要となることが示唆された。
本結果は、学生の卒業論文としてとりまと
められた。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者に
は下線）

〔雑誌論文〕（計 1 件）
井上悦子，吉廣卓哉，中川優，“大規模ク
ラスタリング結果のグラフによるインタラク
ティブな可視化手法，” 電子情報通信学会
和文論文誌 D，査読有，vol. J92-D，no. 3，
pp. 351-360，2009。

6. 研究組織

(1) 研究代表者

井上 悦子 (INOUE ETSUKO)

和歌山大学・システム工学部・助教

研究者番号：50452537

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：