

平成22年06月09日現在

研究種目：若手研究（A）
 研究期間：2006～2008
 課題番号：18680011
 研究課題名（和文） 大規模ネットワークからの目的情報抽出に適した可視化手法の研究
 研究課題名（英文） Visualization for extracting of suitable information from huge network.
 研究代表者
 阿部 浩二（ABE KOJI）
 東京大学 インテリジェント・モデリング・ラボラトリー 特任助教
 研究者番号：80401321

研究成果の概要：利用可能な情報量が膨大になり、目的情報の抽出が難しくなりつつある。そこで、利用者が目にした記事、検索時の行動、動画視聴時の操作、携帯電話からの移動情報等の自動取得・解析を行い、これらの結果を使用した情報抽出支援を行った。情報提供サービスの実装・評価結果から、従来の情報分類指標と異なる情報のさらなる絞込み・利用者の個人特性への対応が可能と考えられる特性指標の取得と、それらを利用し利用者評価の高い情報可視化手法を開発した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	4,500,000	1,350,000	5,850,000
2007年度	3,900,000	1,170,000	5,070,000
2008年度	2,700,000	810,000	3,510,000
総計	11,100,000	3,330,000	14,430,000

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：検索,情報抽出,可視化,ネットワーク,ユーザインタフェース,行動履歴情報

1. 研究開始当初の背景

様々な分野における研究の進展・計算機器の進歩により、人々が利用可能な情報は膨大なものとなりつつある。この傾向は今後ますます強まるものであり、これら大量の情報を持つ余すことなく扱うことを可能にする手法が求められ始めた。現在 Web 等で公開されている各種データベースの多くは強力な検索機能と共に提供され、増え続ける情報の主要な呈示手段の一つと言える。しかしながらこのようなデータベース群は、取得された情報の有効活用に十分なものとはいえず、辞書

的な使用、あるいは統計的な手法の適用先としての利用が大部分を占めている。

このように単純な利用にとどまっている原因として、「新たな知見の発見や情報の持つ本質的な意味を人が考察する際に必要な、情報の持つ背景・他の情報群との関係性の利用が不十分である」、「目的に応じた情報を抽出する手法があまり研究されていない」ことが考えられた。具体例として、遺伝子に関するデータベースを挙げる。前者については、検索時に個別の遺伝子そのものを対象とすることが多く、検索対象を特定・選択した後

によりやく関連する遺伝子の候補や対象遺伝子のもつ背景が確認でき、制御の仕組みの考察などが可能となる点が挙げられる。また後者については、このような要素中心の提示では、得られた複数の情報を人が頭の中で再構成し関係性群としての理解を試みなくてはならず、利用者に多大な負担をかけると共に取得可能な情報群を有効に活用しているとは言い難い状況が挙げられる。これらは生物分野に特有の性質ではなく、類似の問題は情報が蓄積されつつあるほぼ全ての分野で確認されるものである。検索結果のリストを呈示するだけではないインタフェースの研究は、まだ始まったばかりの段階にあると考えられる。

2. 研究の目的

情報から意味を見出す際に重要な「背景」は、「関係性」と言い換えることができる。ある情報をもつ背景とは、同時に対象とされる他の情報から発生するものであり、どのような関係性からその情報をとらえるかということに他ならない。このような視点から、現在高度な理解・活用が望まれている大規模な情報群を巨大な「ネットワーク」として扱い、このネットワークを目的情報抽出のために処理・呈示する手法が必要であると考えられた。

以上の考察から本研究では、大規模情報ネットワークからの目的情報抽出に適した可視化手法の研究を目的とする。具体的には、大規模情報群で構成される関係性ネットワークのモデルの推定、各分野において抽出が望まれる知見と抽出の際に行われる思考プロセスの解析、人が把握可能な大量情報の可視化呈示手法 これらを明らかにすることを試みる。

特定されたモデルは、解析された思考プロセスを定量的に表現する際の基盤となると共に、情報呈示の際のマッピングにおいても利用が可能である。また、解析された思考プロセスは、可視化呈示時において容易に利用可能なフィルタ群として提供することで、現在研究者が所属する分野の視点だけでなく他の分野の研究者による視点からの思考結果を擬似的に取得することが期待される。これらの機能を備えたデータベースの実現には、多くの研究者による利用とその結果のフィードバックが必要とされることから、直観的な操作が可能なインタフェースを持った可視化手法による呈示を検討する。これらの研究から得られた知見を統合し、大量の情報群からの目的情報抽出可能な可視化呈示のための方法論を確立する。

3. 研究の方法

(1) 情報抽出時の思考プロセス解析

関係性ネットワークを構成する情報群に対するフィルタ事例として、情報検索時の研究者による思考プロセスのモデル化がまずは考えられた。利用者への情報呈示の際に、予め解析された個別の利用者による異なる思考プロセスを、可視化呈示時において容易に利用可能なフィルタ群として提供することで、現在研究者が所属する分野の視点だけでなく他の分野の研究者による視点からの思考結果を擬似的に取得可能になると期待された。

まずは、思考プロセスを第三者に提示するインタフェースを実装するための基礎的な知見を得るため、既存のインタフェースを利用した際の情報取得手順について映像で記録し、利用者の意図を推定する実験を行う。

被験者の推定内容は、情報取得手順を記録された研究者自身により再評価がなされ、各操作の意図の把握度合いを5段階で評価することとした。

(2) 動画視聴時の操作履歴情報の解析・利用

人が取り扱う各種デジタル情報の増加に伴い、動画視聴の生活に占める割合も増加し続けている。情報抽出を行う利用者への負担をなるべく増加させず、日常生活における各行動情報の自動取得・解析結果を利用することによる情報抽出支援を試みる。

既存のWeb上でサービス展開している動画情報提供ポータルサイト利用者の6ヶ月分の動画視聴時における操作記録を対象とする。動画視聴時における操作履歴から、動画の特性および利用者の視聴状況により異なると予想された、操作パターンの解析により動画特性の評価・動画分類を行い、動画情報推薦の際に利用可能な特性値の提案を目指す。

(3) テキスト情報閲覧履歴の利用

push型の情報提供を行う際に、利用者の過去のニュース閲覧情報、投稿情報を構造化することにより、単語ベースでの検索ではなく、各語句の使われた前後の文脈性を含んだ情報抽出アルゴリズムの実装・評価を行う。評価方法としては携帯電話を用いた実証実験サービスを実施し、参加者の一部に対するアンケート評価を行い、提供情報の操作することによる提供情報に対する興味度合いの変化を調査する。

(4) 大量情報の可視化提示手法の開発

① 提供情報の可視化手法

情報提供サービスにおいて大量に保持されている提供用情報は膨大であり、検索結果による間接的な理解を除くと、その全容を利用者が把握することは非常に困難である。このため、サービス側が保持している情報群自体を構造化し、その全容を多段階に要約した

グラフとして描画・提供する。テキスト情報を分節しクラスタリングを行うことで構造化は行う。またグラフ描画を行うと同時に、推薦サービスとの関係性を確認できるよう、推薦アルゴリズムで用いられた部分の表示を併せて行うものとする。サービスを実際に利用した人の中から抽出された被験者群に対するアンケートを行い、全容の把握、アルゴリズム理解、サービスの向上への利用といった面からの評価を行う。

②テキスト情報閲覧履歴の可視化手法

各個人に最適化された情報提供を行うために利用者一人一人を区別し保持された情報閲覧履歴を、構造化し各利用者に可視化・呈示する。これにより、通常ブラックボックスとされがちな個人情報の利用の仕方を開示すると同時に、サービス品質を上げるために、個別の行動履歴情報を利用すべきか否か利用者が推薦結果の変化を確認しつつその作業を行えるものとする。評価としてサービス利用者によるアンケート調査により、情報処理速度の向上と、安心・安全に対する意識調査を行う。

③視聴パターン分類された動画提示手法

視聴パターンによる特性値を元に分類された動画群を同時に呈示・選択・閲覧が可能な可視化手法を実装し、呈示される動画が確認できると同時に、その特性理解が可能となる動画呈示手法を開発する。動画の特性値を利用者が入力可能なインタフェースを有すると同時に、言語によるタグのような既存の分類手法との異なる提案特性値による分類が有効であるかを実証実験により検証する。

④行動履歴情報の提示手法

様々なデバイスにより取得可能となってきた行動履歴情報は、活用が期待されているものの、その利用方法については未だ検討段階にあると考えられる。そこで、新たな利用方法創出の支援および、取得されている行動履歴の提供者自身が再確認することによる日常行動への指針を与えるような可視化呈示手法を検討する。具体的には、実感することが難しい、各地域に対する滞在時間を地形図の形状変化として表現し、数ヶ月間に渡る大量の行動履歴のマッピング結果を、利用者が把握可能なものとして呈示する。また、恣意的に決定可能な変形強度を表すパラメータについては利用者が変更可能なものとし、それらの数値取得により、数量的な時間と体感的な時間を対応させる際の適切な数値取得およびその数値の個人差、分類傾向といったものを調査・取得する。

4. 研究成果

(1) 情報抽出時の思考プロセス解析

被験者の推定内容は、情報取得手順を記録された研究者により各操作の意図の把握合いを5段階で評価した。この結果から、通常のWebブラウザを用いた既存のインタフェースであっても、閲覧滞留時間・マウスカーソルの移動・検索手順などから、情報取得の目的及び情報検索者の専門分野の推定は70%前後の精度で可能であることが判明した。その一方で、情報検索者のバックグラウンドとして重要であると考えられる、研究者の年齢・所持していると思われる関連知識等についての推定結果には、大きな間違いも多く見受けられ、各個人を特定するようなレベルでの意図をの把握は非常に困難であることが明らかとなった。

このため、研究に長く携わっている研究者の人的資源をより活用するために、その知識体系を明示的に提供することによる意図の把握補助が有効であることが明らかとなった。このような支援のため、各研究者による過去のテキスト表現を収集・構造化し、情報検索時の重み付けフィルタとして利用可能とすることが有効であることを示すことが確認された。

(2) 動画視聴時の操作履歴情報の解析・利用

既存の分類とは異なる、「その動画への人の接し方」という観点からの動画評価の実現を目指した。動画視聴時の操作ログの履歴を利用し一連の行動の意図・目的を分類、推薦時に利用可能とした。具体的には「動画共有サービス利用時の操作履歴情報を解析、利用者の行動パターン分類を抽出」、「動画・利用者共に上記分類による特性値によるラベル化・分類」、「動画と利用者の適合度は、特性ラベル値を成分とするベクトルの内積値で評価・推薦」である。

行動パターンの抽出に使用した操作履歴情報は、動画共有サービス：サグールテレビ(チームラボ株式会社)において2008.03.27~2008.09.30に取得された、各動画の再生・シーク・スキップ・終了に関する情報である。本事業では操作履歴を対象とするため、何かしらの操作履歴が存在し総再生数が30回を越える動画17,337本について、個別に中身を解析・確認していく対象とした。4種類(再生・シーク・スキップ・終了)の情報で構成される操作履歴情報に対し、目的が想定される5種類の行動パターンモデルを仮定し、各パターン特性の現れ度合いについて、0~1で正規化し、動画の特性情報とした。分類後の実在動画群に対し視認による確認を行ったところ、各パターンの代表例として想定された動画が一定数以上含まれていることが確認された。パターンの種類・閾値により特性の確認具合は変動するものの、動画推薦時

に利用可能な特性値として利用可能であると考えられた。さらなる方向性としては、これら閾値設定の自動化、パターン抽出・規則化の自動化が課題として考えられた。

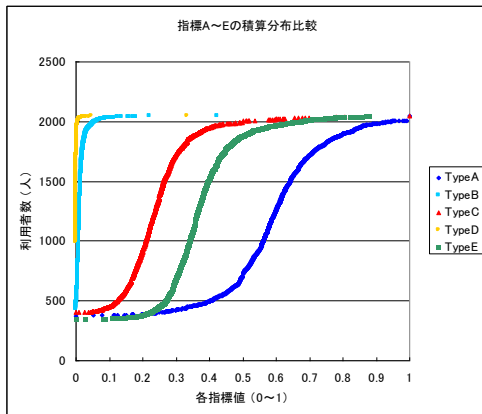


図 4.1 利用者群中における使用指標分布

(3) テキスト情報閲覧履歴の利用

情報閲覧履歴情報の公開に対し、同意を得られたサービス利用者に対し、アンケート調査を行った。行動履歴蓄積期間を変えた2群に対し、文脈製を加味する拡張を行った場合・行わなかった場合それぞれの組み合わせ4通りについて、各パターン約200件の推薦候補情報を設定した。これらの情報呈示結果に対する評価からは、既に見た・知っている情報の出現頻度が下がり、詳細を見たい・興味があるといった反応を増加させることが確認された。その一方で履歴の蓄積に伴い、徐々に精度が上がるのではなく、提供される情報傾向の偏りが強化されてしまうことも一部指摘された。

これらの結果から、情報閲覧履歴を用いることにより、利用者により適合した情報が提供できることが確認された。今後の課題としては最適化され過ぎた結果、提供情報がある特定の狭い範囲に嵌りこむことによる弊害を避ける手法の実現が期待される。

(4) 大量情報の可視化提示手法の開発

①提供情報の可視化手法

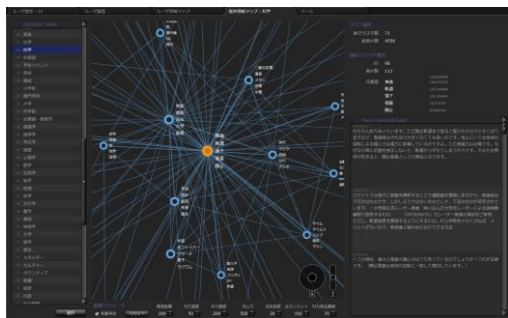


図 4.2 提供情報分類提示(部分)

図 4.2 は利用者に表示された提供情報のクラス

タリング結果である。このような可視化呈示結果により、多くの分野の情報が含まれる情報群の各分野でどのような関係性のある情報が提供されているのかを確認でき、利用可能な情報を確認できるものとした。

また、各利用者への情報推薦根拠となった該当部分の表示、各利用者の情報閲覧履歴との関係を併せて表示することにより、サービスを利用する上で使用されてほしくない履歴情報の取捨選択が高速に行えることが確認された。

アンケート調査からは、情報提供サービスがある程度受け続けることでしかそのサービスの特性が把握できない場合に比べ、サービス自体の利用方法も容易に想定できることは有効であるという利用者観点からの評価が得られた。また情報提供者側の観点からの、推薦処理によりどのような情報が提供されることになるのか把握可能であるため、サービス提供時に、より効果的な提供すべき情報の選別が可能となるといった評価が得られた。

②テキスト情報閲覧履歴の可視化手法



図 4.3 利用者への推薦情報履歴表示

提供情報とマッチングを行う、利用者側の情報閲覧履歴についても可視化呈示を行った。本研究での実証実験サービスでは、位置情報を情報提供の際にフィルタとして用いたため、実空間情報である2次元地図上へのマッピングを試みた。地図上の各点一つ一つが、利用者が推薦を受け閲覧した情報履歴である。このような履歴呈示をすることで、利用者はどの地域に関する情報を自身が求めていたのかを確認することが出来ると同時に、情報提供サービス側に利用された個人情報の内容を把握することが可能となる。このような情報呈示に対する評価として、漠然と個人情報の提供への抵抗感があつた利用者層から、個人情報の使用内容が明確になることにより、不安感が減少するとの評価を得ることができた。この結果は、情報漏洩を防ぐ方向性とは異なる手法による安心・安全性の提供の可能性を示したと考えられた。

③視聴パターン分類された動画提示手法

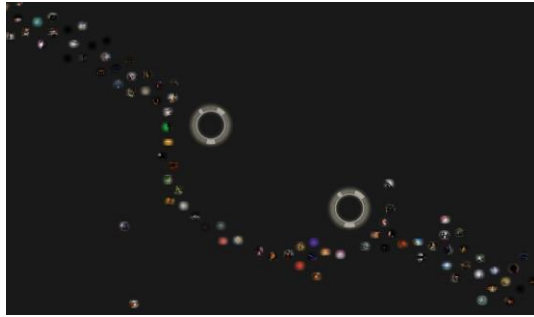


図 4.4 動画分類結果呈示例

設定した特性値による動画分類・呈示例が図 4.4 である。図中の白いリングに見えるものが各特性値を比率で持つ一利用者プロフィールをあらわしており、類似度により周辺に推薦動画が配置される。ズームイン・アウトが連続的に行え、個別の動画視聴もシームレスに行うことが可能である。



図 4.5 動画分類結果呈示拡大例

個別の動画は図 4.5 のように拡大表示が可能であり、また各種動画に割り振られた特性値の確認、分類特性の変更が行える。

このような可視化呈示手法を実装し、利用者による評価を行った。その結果からは、「視聴パターンとして設定された特性値そのものの内容を理解することは難しいが、同じテンポで動画を見ていくようなことが出来る。」あるいは、「異なる分野の動画も混じっていたが、同様に楽しめる動画が次々繋がっており面白かった。」「日常的に使用してみたい。」といった評価が得られた。これらの結果からは、動画が予めタグ付けられた分野としての分類とは異なる、視聴スタイルによる動画評価結果が、動画情報の分類・推薦に利用可能であることが明らかとなった。従来使用されている分類タグとは独立した性質

のものであるため、両者の併用が可能であり、組み合わせによる高度な絞込みへ今後利用が期待される。

④行動履歴情報の提示手法



図 4.6 行動履歴情報呈示例

携帯電話を利用し取得された行動履歴情報の呈示手法として、図 4.6 に示されるようなインターフェースを開発した。移動軌跡に対する色分けは、それぞれ移動が行われた時間帯を示しており、活動時間帯が重なる地域は（色の重ね合わせにより）より白色に近づいていく。また、滞在時間が長い、あるいは訪問頻度の高い地域は、地形図の変形により拡大され、かつ、より細かい道路情報等が把握できる縮尺の地図へと置き換えられる。

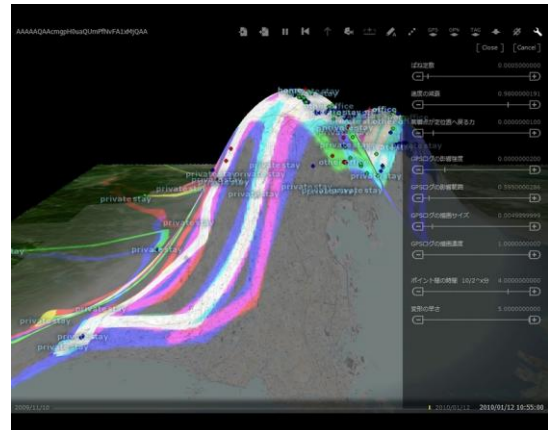


図 4.7 地形変形用パラメータ設定画面

行動結果の地形変形に寄与する強度は利用者により任意に変更することができ、2D⇔3Dの切り替えも可能である。

このような提示手法に対し、3ヶ月間取得された各利用者の行動履歴情報を利用者自身により閲覧してもらい、行動履歴情報を可視化呈示されることについての評価を取得した。

アンケート結果からは、92%の利用者が、滞在時間に応じた地図の拡大・縮小表現による情報呈示を効果があり体感に沿うものであると考えたことが分かった。

自由回答内容からは、「わずか3ヶ月と思っていたが忘れていた行動も多く、それらが思い出された。」「情報蓄積が長期化すると、より効果を発揮しそう。」といった意見も得られた。

これらの結果からは、実際に移動した場所を正確に呈示されることは、普段頻繁に訪れている気になっている地域のうち、実はほんの一部へしか行っていないことを利用者に気付かせる効果があることが分かると同時に、新たな行動誘起の際の候補地や、情報抽出・提供の際へ行動履歴情報の利用可能性が示せたと考えられる。

5. 主な発表論文等

〔雑誌論文〕(計1件)

“阿部浩二, 廣瀬通孝” 大量情報群からの目的情報抽出に適した可視化手法に関する研究, 日本機械学会 第16回 設計工学・システム部門講演会講演論文集, 2006, p. 345-346, 査読無

〔学会発表〕(計1件)

“林織部, 西村邦裕, 阿部浩二, 谷川智洋, 廣瀬通孝” 大量動画視聴における注意誘導に関する研究, 電子情報通信学会 技術研究報告, 2009年6月9日, 東京大学

6. 研究組織

(1) 研究代表者

阿部 浩二 (ABE KOJI)

研究者番号: 80401321

