

研究分野と参照関係からたどる文献可視化インタフェース

中澤 里奈 伊藤 貴之 斎藤 隆文*

概要. 研究を進めていく上で、論文サーベイは非常に重要である。しかし、初学者は検索する上で適切なキーワードを全て知っているわけではないため、検索エンジンを用いた方法では、検索ワードによっては見落としがおこる可能性がある。一方で、より視覚的に参照関係をたどりながら文献を調査する方法の一つとして、文献の参照関係の可視化の研究が数多く行われてきた。これらの研究では、文献をノードとして時系列に並べ、参照関係をエッジとしたグラフによって、参照関係の構造を表現している。このような手法は、研究分野をまたいだ複雑な接続関係が視認しづらく、参考文献をたどっていくにはやや不向きである。そこで本報告では、研究分野と参照関係からたどる論文探し支援のための参照関係可視化手法を提案する。提案手法では、文献の内容をもとに文献を分類し、これらの分類と参照関係にもとに、可視化手法 FRUITS Net を用いて文献の配置を決定、エッジバンドリングを用いて参照関係情報の要約可視化を行う。これにより、分野間の参照関係をより簡単に把握することで文献調査を支援できるのではないかと考える。

1 研究背景

研究を進めていく際の文献調査は非常に重要な要素である。調査方法としては、検索エンジンを用いたり、既に読んでいる論文の参考文献をたどっていったりするといった方法がある。しかし、研究を始めたばかりの学生は検索のための適切なキーワードを全て知っているわけではないため、検索エンジンによる方法では、検索ワードによって見落としがおこってしまう上に、論文同士の関係を見いだすことができない。一方で、参考文献をたどることで視覚的に文献を調査する方法の一つとして、文献の参照関係の可視化の研究が数多く行われてきた [1] [2]。これらの研究では、各文献をノードとして時系列に並べ、参照関係をエッジとしたグラフによって、参照関係の構造を表現している。このような手法では、研究分野をまたいだ複雑な参照関係をグラフとして描画すると、非常に多くのエッジ間の交差が生じて視認性に悪影響をおよぼす。そこで本報告では、研究分野と参照関係からの文献探しに適した参照関係可視化手法を提案する。本手法では、文献の内容でクラスタリングされた参照関係グラフを描画し、クラスタ間の参照が多い場合にはエッジバンドリングを施す。この時、敢えて時系列順に並べないことでクラスタ間の参照関係の太さを視認しやすくする。これにより、一つのキーワードに関係ある論文が複数の分野にわたっている時に、それらを見逃さずに網羅的に発見でき、その全体的な構図も理解しやすくなると考えられる。

2 提案手法

2.1 文献のカテゴリ分類

本手法では、文献の内容から文献を研究の分野別にカテゴリ分類するために、今回は一例として潜在的ディリクレ配分法 (LDA: Latent Dirichlet Allocation) [4] を用いた。LDA は、一つの文書に対して複数のトピックの存在を想定した、文書生成過程の確率的なトピックモデルである。LDA で推定されるトピックを一つの研究分野と見なし、各文献がどのような研究分野に属するかを分類する。LDA において、トピック数は予め決めておく値であるため、今回はトピック数を 10 とした。

2.2 参照関係の要約可視化

続いて、LDA を用いて推定したカテゴリ分類をもとに、FRUITS Net[3] の配置アルゴリズムを用いることでノードの配置を決定する。これにより、参照関係がある文献同士を近く、同じ研究分野に属する文献同士を近く配置し、ノード同士の重なりを回避しつつ、画面配置占有面積を低減することができる。ノードの大きさはその文献の被引用数に比例する。その上で、エッジの視認性を向上するためにエッジバンドリングを施すことで、大量の線情報を束ね、要約する。エッジを束ねる本数はユーザが閾値をスライダによって設定することができる。カテゴリ分類によってできた文献クラスタ間のエッジ本数を計算し、それが閾値以上であるならば、本数に比例させた太線を束として描く。閾値に満たない場合はエッジをそのまま描画する。

2.3 ユーザインタフェース上での描画と操作

図 1 は、本ユーザインタフェースにおける表示画面例である。ユーザは操作パネル上の (1)(3) で

Copyright is held by the author(s).

* Rina Nakazawa and Takayuki Itoh, お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻, Takafumi Saito, 東京農工大学大学院 工学研究院 先端情報科学部門

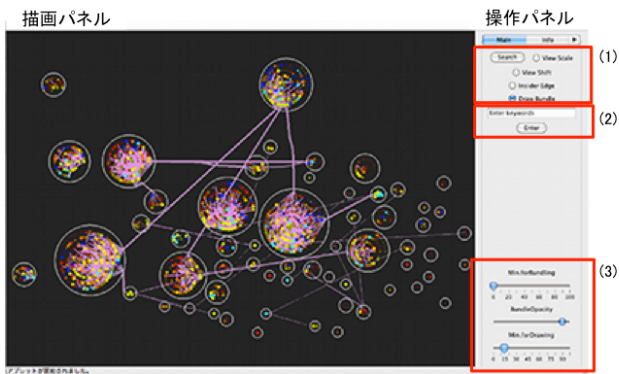


図 1. インタフェース表示画面例

可視化結果の拡大縮小や平行移動操作，エッジバンドリング適用の有無，その閾値の設定を行うことができる。また，各文献であるノードをクリックすると，その文献のタイトルや発行年，アブストラクトなどの詳細情報を表示する。同時に，クリックされたノードが持つエッジと，それに接続するノードが持つエッジの二世代分の参照関係をハイライト表示する。このエッジのハイライトは2つのノードについて同時に適用可能であり，これによって両ノードの接続関係を比較することができる。一方，最初に注目すべき文献を探すことは初学者にとっては容易ではない。そこでユーザはだまかな研究分野かキーワードを利用することで，画面上に表示させる文献をフィルタリングすることができる。自分の興味のある研究分野カテゴリを操作パネル上のボタンから選択した場合，選択した研究分野のみと分類された文献群をウィンドウ中央部に拡大表示される。また，(2)にキーワードを入力すると，入力された語がタイトル中に含まれている文献ノードのみが表示される。

3 実行結果

使用した文献データは1990から94年，2000から2010年までにSIGGRAPHで発表された論文1072本である。例として，ユーザが写真や動画の処理関連の研究の文献を調査したい場合を考える。写真，動画関連の研究分野カテゴリを選択すると，表示結果は図2のようになる。図2の(A)が写真，動画の処理関連の分野であり，(A)に接続する束を追っていくと，(A)には(B)と(C)のどちらかと参照関係を持つ文献が多く存在するということがわかる。(B)は画像処理関連のクラスタ，(C)は大域照明，光反射や光拡散モデルなどと写真や動画の処理が組合わさったクラスタである。この結果から，写真や動画の処理に関する文献を探したい場合，優先して探したい内容がどちらかという(B)の画像処理技術に関連寄りなのか(C)のライティング関連寄りなのかによって，その分野と接続している文献から先に

読み進めていけばいいということがわかる。従って，本インタフェースを利用することで関係ありそうな論文を研究分野と参照関係から整理して，どの論文群から先に読むかを選択できるのではないかと考えられる。

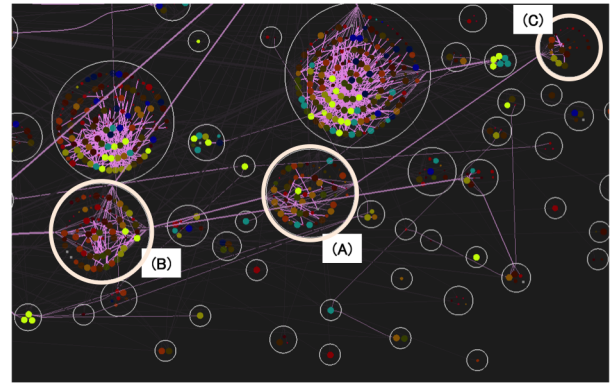


図 2. カテゴリ選択時の表示結果

4 まとめと今後の課題

本報告では分野間の繋がりを考慮した文献参照関係の可視化インタフェースを提案した。本手法によって，研究分野と参照関係から関連のありそうな文献を整理し，優先的に読むべき文献を選ぶことができるようになるのではないかと考えられる。現在はエッジの方向性を考慮した描画を行っていないため，今後はエッジの方向性を表現したエッジバンドリングと可視化を行っていく。さらに，今回手動で決定していたトピック数の推定や GUI 機能の拡充，評価実験なども進めていきたい。

参考文献

- [1] J. Matejka, T. Grossman, and G. Fitzmaurice. Citeology: visualizing paper genealogy. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 181-190, 2012.
- [2] 岡田, 渡部, 斎藤. 時系列に基づいた文献参照関係の可視化-文献のクラスタリング-. 情報処理学会第74回全国大会, 3ZB-3, 2012.
- [3] T. Itoh, C. Muelder, K. -L. Ma, and J. Sese. A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 121-128, 2009.
- [4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.