

文献の関係情報視覚化による研究領域サーベイ支援ツール

朴 春子[†] 三末 和男[†] 田中 二郎[†]

[†]筑波大学大学院コンピュータサイエンス専攻

1 はじめに

新しい分野の研究をはじめるとき、サーベイによって研究領域の発展をいち早く把握する事が重要だと考えられる。研究領域を把握する有力な手法として可視化が挙げられる[1, 2]。ただし、既存の手法はキーワードと著者の関係、引用関係だけの分析を行っているものが多い。

研究領域を把握するためには対象領域の基本となる文献と、自分の研究テーマの関連研究を探ることが重要である。手に入れた一つの文献から、キーワードによって同じテーマを扱っている文献を探ることができる。しかし、テーマ内容が似ていても文献の発表年によって、新興分野では既存の分野と異なるキーワードを使っている場合がある。したがって、関連文献を探す時、文献間の内容的な関係を表す引用関係の可視化は重要である。さらに、文献のキーワードと引用関係を連携することによって、対象領域の重要文献と関連文献の探索を支援する事ができる。

本研究では、対象領域の文献のキーワードと引用関係が一覧できるツールの開発を目指す。アプローチとして、対象領域のキーワードの提示と、文献の重要度を測る四つの指標が一覧できる可視化手法を設計し、実装する。

2 文献の重要度を測る指標

文献検索を支援するために、文献の重要度を測る参考情報として、四種類の情報を提示する。

1. 発表年

最近の文献を読むほうが最近の研究動向を把握でき、新しい研究点を見つけることができる。

2. 全体被引用数 (引用している文献の全部の数)

全体分野における被引用数が高い文献ほど

基本となる重要な文献として研究者によく使われる。

3. 特定被引用数 (引用している文献で着目した分野に属する文献の数)

研究者が着目した分野内の文献からよく引用されているなら、着目した分野内では重要な文献であると考えられる。

4. キーワード

選択した関連キーワードを含む数が多いほど、研究者の関心があるテーマとの関係が強いと考えられる。

3 ツールの開発

キーワードと引用関係を連携するために、キーワードを表現する「タグクラウド」(図1下部)と、文献間の引用関係を表す「引用関係図」(図1上部)を連携したツールを開発した。

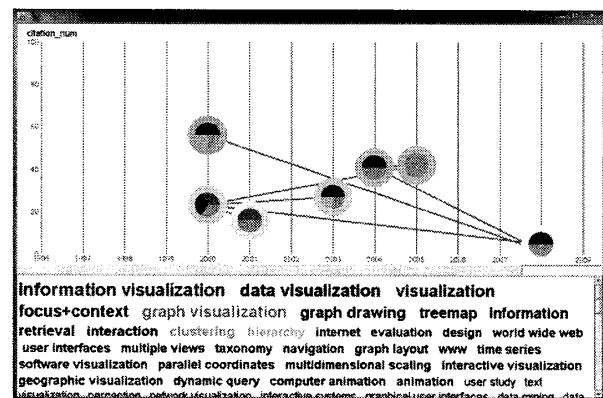


図1. ツールの概観

3.1 キーワードのタグクラウド

キーワードをアルファベット順と使用頻度数の両方の観点から迅速に特定できるようにタグクラウドを採用した。タグクラウドの各タグはキーワードを表し、タグのサイズはキーワードの利用頻度(キーワードを含む文献の数)を表す。さらに、選択したタグに異なる色を付けて表示する。また、キーワードの検索を便利にするために、タグクラウドの右上に入力バーを付

Tool for scientific domain survey by visualizing literature relationship

[†] Chunzi Piao, Kazuo Misue and Jiro Tanaka

[†] Department of Computer Science, University of Tsukuba

けた。

3.2 文献引用関係図の描画

文献の四つの指標を表すために、各文献を以下の視覚化表現を持った円グラフで視覚化して図 1 のように描画した。

文献の発表年を X 軸によって表した。最近の文献ほど右に配置される。

文献の全体被引用数を Y 軸によって表した。全体的被引用数が高い文献ほどツールの上に配置される。

文献の特定被引用数を円の外側のリングの色の濃さで表わす。特定被引用数が高い文献ほど濃い色で表わされる。

文献が含むキーワードを対応する色で分けた円で表わす。円の色は文献が含むキーワード (タグクラウドから選択したキーワード) の数に対応する。

文献間の引用関係を視覚化するために、引用関係を線分で表現する。円と円をつなぐ線分がある時、右の文献が左の文献を引用していることを示す。

4 ケーススタディ

ここでは、研究者がクラスタリング研究に興味をもっていると仮定する。ツールの初めの画面で、研究者が「clustering」を選択するとタグクラウドに clustering を含む文献と、それらの引用文献及び被引用文献に使われた全部のキーワードが使用頻度によって表示される。

タグクラウドの関連キーワードから自分が興味を持っているキーワードを複数選択すると、選択したキーワードに異なる色がつき、図 2 のように更新される。

描画結果の引用関係図 (図 2 上部) を俯瞰することによって全体被引用数が一番高い文献 D を見つけることができる。さらに、引用関係を見ることで同じキーワードを含んでないが内容的に関連がある最近の文献 A も見つけることができる。それによって、最近の研究発展の方向を把握する事ができる。

また、特定被引用数が高い文献 C を見つけた後に、それを引用している最近の文献 A を見つけることもできる。それによって、最近の関連文献を見つけることができる。

また、発表年とキーワードの二つの観点から

重要文献を探索する事もできる。例えば、最近の文献の中で選択したキーワードを多く含む文献 B を見つけることができる。

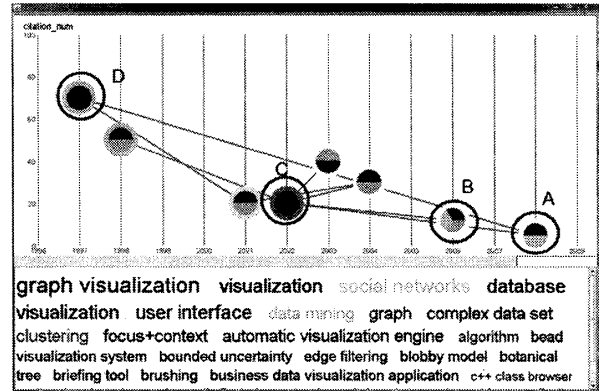


図 2. 描画例

5 関連研究

文献情報可視化として Chen[1]は文献の引用年、引用数を視覚化して最近の研究トピックを分析している。Ke ら[2]は、発表年、引用数、引用年による論文の引用関係から研究分野に影響を与えた論文などを発見できることを示した。本研究では、引用関係とキーワードを連携していることが特徴である。

6 まとめ

サーベイを支援するためにキーワードを表すタグクラウドと文献の引用関係図を連携するツールを開発した。さらに、文献探索に役に立つ情報として、文献重要度を測る参考値となる四つの情報を連携する可視化手法を提示した。

参考文献

[1] Chaomei Chen. The Centrality of Pivotal Points in the Evolution of Scientific Networks, Proceedings of the 10th international conference on Intelligent user interfaces (IUI '05), pp. 98-105, 2005.

[2] Weimao Ke, Katy Borrner, Lalitha Viswanath. Major Information Visualization Authors, Papers and Topics in the ACM Library, Proceedings of IEEE Symposium on Information Visualization(InfoVis '04), pp.r1-r2, 2004.