

文献参照関係の可視化 – ノード配置の改善

村松 千春[†] 斎藤 隆文[‡]

[†]東京農工大学 工学部情報工学部

[‡]東京農工大学 大学院生物システム応用科学府

1 背景と目的

研究を行う際に、その研究と同じ分野の文献を調査することは重要である。また、ある文献の参照を辿ることによって、多くの研究の基礎となる研究や、その分野において注目される論文を探すことが可能である。しかし、大量の文献の中から目的の情報を探し出すことは困難である。そこで、文献の参照関係を図示することで、ある分野の先駆けとなる文献、注目を集めている文献などの発見が期待できる[1]。本研究では、文献数が膨大であっても参照関係を容易に把握でき、文献の関係性や重要度などが見て取れる、時系列に基づいた文献参照関係の可視化手法を目指す。なお、本稿では先行研究によって行われた可視化の結果を、より分かり易く表示する方法を提案する。

2 文献参照関係の可視化

岡田らにより提案された文献参照関係の可視化手法[1]について説明する。

2.1 文献データ

使用する文献データは ACM Digital Library[2]から取得している。同サイトでは著者、文献名、発表年、掲載誌などの情報の他に、その文献が参照している、または参照されている文献についても記載されている。これらから必要な情報を抽出し、可視化のデータとして使用する。

2.2 可視化手法の概要

可視化は図 1 に示すようにして行う。各ノードは文献を表し、図 2 に示す疑似カラーを用いて文献が参照された回数を表現する。ノードは、横軸は年代順に、縦軸をクラスタリングによって割り当てられたクラスタによって割り当てられたクラスタに対応した場所に設置する。

リンクは文献の参照関係を表わす。参照の場合はノードの左側から、被参照の場合は右側からリンクを描画する。可視化の結果は図 3 のようになる。

2.3 クラスタリング

各文献に対して参照関係に基づいた類似度を設定し、階層型クラスタリングを行う。

類似度は、以下の 3 つの条件より算出する。

- ・直接的な参照・被参照関係が存在
- ・間接的な参照・被参照関係が存在
- ・共通の参照・被参照関係を持つ文献が存在

それぞれの条件を点数として加算することで類似度を算出し、値が高いほど 2 つの文献は類似していると考ええる。

また、木構造を作成する際に参照回数の少ないノードを除いておくことで、小さなクラスタが生まれない。この除かれたノードは、非階層型クラスタリングを行い、最も類似度の高いクラスタに所属させる。

2.4 リンクの描画方法

この研究では、膨大な量の文献を可視化することを目的としている。二次元上に単色で膨大な量の文献へのリンクを描画した場合、どのノードからリンクが出入りしているのか識別することができない。そこで、図 4 で示すように z バッファを利用し、グラデーションを用いてリンクの描画を行う。ノード付近では手前にリンクを描画し、ノードから離れると奥にリンクを描画する。

2.5 フィルタリング

リンクの描画方法を工夫することで、どのノードにリンクが出入りしているのかの識別が可能になったが、リンク数が多い場合、各リンクがどのノードと繋がっているかまでは判断できない。そこで、ノードを選択し、そのノードと参照・被参照関係のあるノードとそれに接続されるリンクだけを表示するフィルタリング機能を提供する。

Visualization of document reference relations - Improvement of node placement

Chiharu MURAMATSU[†], Takafumi SAITO[‡]

[†]Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

[‡]Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology.

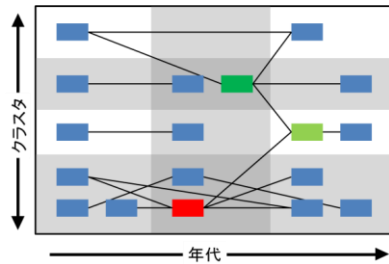


図 1. 可視化概要



図 2. 疑似カラー

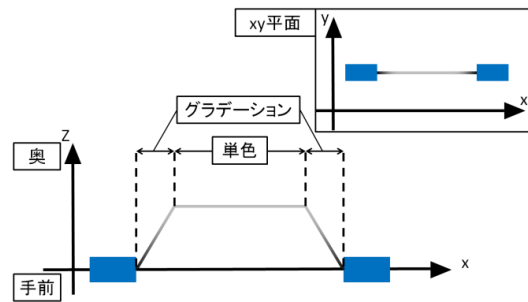


図 4. リンクの描画方法

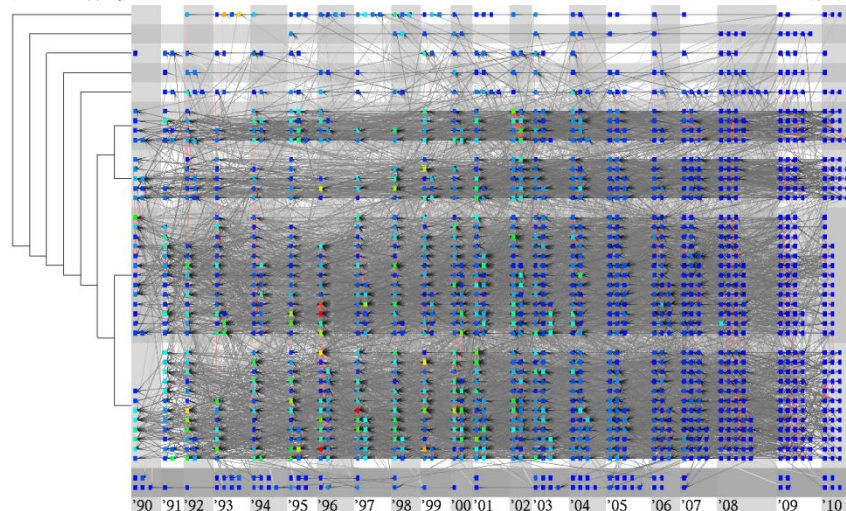


図 3. 可視化結果

3 ノード配置の改善

2 節で説明した文献の参照関係に基づく可視化手法により、参照関係が集中している文献や、対象の文献と間接的に参照関係のある文献などの文献同士の関係性を調べることが容易となった。しかし、フィルタリングを使用せずに全体を見つつ、同時に個々のリンクを識別することは困難である。そのため、可視化結果の観察を容易にする手法が必要であると考えられる。

クラスターに分類した後のクラスター内でのノードの位置については、ACM Digital Library でデータを取得した際の ID に従っている。ここで参照関係に配慮しないノード配置を行っていることで、リンク同士の交差が多発し、個々のリンクの識別が困難になるのではないかと考える。

この問題を解決するため、対象文献と参照関係のある文献がどのクラスターに所属するか注目し、自身より上のクラスターとの参照回数が多い文献をクラスター内の上に、自身よりも下のクラスターとの参照回数が多い文献をクラスター内の下に配置するように変更する。その結果、クラスター内の文献同士でのリンクの交差を減らすこ

とが出来た。これによりリンクの識別が簡単に行うことが出来るようになったと言える。

4 おわりに

本研究では、時系列に基づいた文献参照関係の可視化を行い、その結果をより理解しやすくするための手法を提案した。

クラスター内の機械的に配置されていたノードを、参照関係をもとに並び替えることによって、クラスター内のノード同士のリンクの交差が減少した。これにより、個々のリンクが識別しやすくなったと考えられる。

今後の課題として、クラスターの配置変更によるリンクの交差の削減、ノード・リンクの削減、クラスターリングに利用する類似度の妥当性の検証が挙げられる。

参考文献

- [1] 岡田拓也, 渡部秀文, 斎藤隆文: 時系列に基づいた文献参照関係の可視化, 第 74 回情報処理学会全国大会, 3ZB-3 (2012).
- [2] ACM Digital Library, <http://portal.acm.org/dl.cfm>