

論文内容の要旨

博士論文題目 Link Analysis with Kernel Metrics
(カーネル距離を用いたリンク解析)

氏名 伊藤敬彦

(論文内容の要旨)

引用関係から文書間の関係を推定する研究が計量書誌学でなされてきた。近年、これらの研究はWebページの持つハイパーリンクの構造を分析するために応用されている。多くの研究がなされている中で、大半の研究が引用関係上に定義される尺度(引用解析尺度: 重要度, 関連度)に関するものである。これに対し本論文は、グラフ上に定義されたカーネルをリンク解析に適用し、カーネルと既存の引用解析尺度との関係について議論している。その中で、グラフカーネルの一つは関連度と重要度を二つの端点とする線分上の任意の midpoint として定義され、重要度もしくは関連度に対する偏りはパラメタによってコントロールされることが示される。

本論文のもう一つの目的は、カーネルを用いて、共引用解析、書誌結合と呼ばれる二つの古典的な関連度算出法の欠点を克服することである。例えば、これらの古典的手法では二つの文書が同一の文書を引用する、もしくは、同一の文書から引用される関係にない場合、それらの間に関連度を与えることができない。この欠点がグラフのラプラシアンに基づくカーネルを用いることで解決されることを示す。

さらに、カーネルを複数のコミュニティが存在する引用関係グラフに適用する際に生ずる問題点について議論する。カーネルに基づくリンク解析はパラメタ調整によって、個々の論文に対して、関連論文あるいは重要論文をランク付けして提示できる。しかし、パラメタを重要度に偏らせると、個々の論文が属するコミュニティにかかわらず引用グラフ全体における支配的なコミュニティの重要論文のみが上位にランキングされてしまう。本論文では、引用の生成過程をモデル化し、その結果を用いて、コミュニティごとの重み付き引用グラフ(コミュニティグラフ)を生成する。これらグラフは、引用の各トピックに対する帰属確率を計算、弧の重みとするため、同一の引用が複数のコミュニティグラフに、異なった重みのもとで存在することができる(多重コミュニティ)。これらのコミュニティグラフにカーネルを適用することによって、対象論文が属するコミュニティを考慮して重要論文を推薦、提示できることを示す。

(論文審査結果の要旨)

平成 18 年 12 月 27 日に開催した公聴会の結果を参考に平成 19 年 2 月 14 日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動が続けていくための十分な素養を備えていることを示すものと認める。

伊藤敬彦は、本博士論文において、カーネル関数を用いた尺度に基づいて論文間の関連度や重要度を評価する手法を提案、および、その意味付けを明らかにし、論文間の引用関係を対象としたリンク解析手法に新たな領域を開拓した。本論文の貢献は次のようにまとめることができる。

- 共引用解析や書誌結合などの古典的な関連度算出法では定義できなかった論文間の関連度や重要度に対して、カーネル関数に基づく数値尺度を利用することにより、関連度・重要度に対する新たな解釈を提案したこと。また、ラプラシアンに基づくカーネルによって、これら二つの尺度を統合した尺度が定義できることを示したこと。
- 複数のコミュニティが存在する場合の引用関係グラフに生じる不具合について検討を行い、各節点がそれぞれのコミュニティに所属する確率を定義することによって、引用関係グラフ全体を複数のコミュニティの重ね合わせとして定義し、これにカーネルを適用することによって、論文が所属するコミュニティを考慮した関連度と重要度を定義できる枠組みを提案したこと。

論文間の引用関係に基づくリンク解析のために、カーネル関数による尺度の利用を提案し、関連度と重要度を測るための新たな手法と意味づけを与え、また、複数のコミュニティが存在する場合の引用関係グラフの解析法を提案することによって、より詳細な関連度と重要度に関するランク付けを行なう方法を提案した本研究は、独創性が高く、有用であり、リンク解析において高い貢献があると評価する。

よって、本論文は、博士（工学）の学位論文として価値あるものと認める。