研究室における研究論文の関連研究推薦システムの開発のためのクラスタリング手法

2016/1/23

j148015　宮下弓槻

0.前回発表での意見

この分野の最新の研究は終わっている？

　→　クラスタリング手法などは、最新のもある。

リサーチクエスチョンがない

　→検討中

立ち位置は？

　→研究室内に限定する。

1.研究の目的・目標

　 研究者は研究の課程において、まず、自分と似た研究分野の論文を集め、読まなければならない。また近年、学術情報の爆発的な増加とともに、論文が電子化されたため、論文自体を手にいれることは容易になった[1]。しかし、論文や文献を検索する際には、タイトルや著者名等の限られた特定の検索キーを利用して検索作業を進めなければならなく、利用者が求めている情報とは異なる多くの不要情報が利用者に知らされてしまうばかりか、目的とする内容の論文が見つけられない等の不都合があった[2]。そこで、目的とする論文を容易に検索できるように、それらをトピックごとに整理・組織化しデータベース化する=電子図書館を構築する研究の必要性が近年益々高まってきており、そのような研究が多く為されている[1]。

　本研究では、研究室初学者を対象とし、研究室初学者はまず、自分の研究室で行われている研究を調査することから研究活動を行うと考えられる。また、研究室初学者は、研究に関する知識も不足しているので、自分の研究分野と同じ論文を見つけるのも難しい。なので、研究室内の卒業論文、修士論文のデータベースから自分の研究分野の論文を推薦してくれる関連研究推薦システムの開発のためのクラスタリング手法を提案する。

2.関連研究

論文間の関係を分析する方法に、書誌結合、共引用分析が古くから知られている。

書誌結合とは、論文の関連度を測る時に、2論文間でどれだけ同じ論文を引用しているか、という基準に基づいている手法であり、難波ら[4]は被参照論文の参照の理由を考慮した参照構造を用いて論文間の類似度を測る手法を提案した。

また、共引用分析とは、2つの論文でどれだけ他の論文に共に引用されているか、という基準に基づいた手法である。ある論文に引用された複数の論文は互いに主題を扱っているという理論であり、これも論文間の類似している度合いを知ることが可能である。この値が高いほど2つの論文が類似しているという尺度になる。

しかし、書誌結合、共引用分析等は論文間の関係の存在を調べるためには有用であるがクラスタリングするためには情報が不足している。

吉田ら[2]は、論文間の関係を分析するために、大量のデータの中から相関関係、規則性、パターンを取り出すデータマイニングの手法であるアソシエーションルール発見アルごリズムを適用することにより論文間の関係を分析し、研究の流れを抽出し、得られた関係を利用して論文をクラスタリングすることにより、研究のマクロな流れを表現する手法を提案した。しかし、新しい論文は参照されている回数が少ないため、吉田らの手法では、新しい論文を含むアソシエーションルールを発見できていない。

　榊ら[1]は、従来の2種類のアプローチ(引用関係を基にした手法、内容を手がかりとした手法)を用いた、より確実性の高い論文の組織化手法を提案した。

←本研究で改善？

3.研究のアプローチ

　クラスタリング手法を研究する。

4.今後のスケジュール

2016年1月～2017年2月：関連研究の分析、クラスタリング手法の分析

2017年3月：卒論のテーマ決定・研究室内論文のデータベース構築

2017年4月：研究を進める

参考文献

[1] 榊剛史,松尾豊,市瀬龍太郎,武田英明,石塚満:論文データベースからの研究トピック抽出,人工知能学会全国大会(第19回)論文集(2005)

[2] 吉田誠,小林隆志,難波英嗣,奥村学,横田治夫:"Research Mining:研究論文データベースからの研究のマクロな流れの抽出"第14回電子情報通信学会データ工学ワークショップ(DEWS2003)論文集.No.7-P-1.Mar.(2003)

[3] 難波英嗣,神門典子,奥村学:論文間の参照情報を考慮した関連論文の組織化、情報処理学会論文誌Vol.42,No.11,pp.2640-2649,(2001)

[4] 難波英嗣,奥村学:論文間の参照情報を考慮したサーベイ論文作成支援システムの開発,自然言語処理Vol.6,No.5,P43-62,(1999)