

|              |                                                                               |
|--------------|-------------------------------------------------------------------------------|
| Title        | 論文間の参照情報の抽出と利用に関する研究                                                          |
| Author(s)    | 難波, 英嗣                                                                        |
| Citation     |                                                                               |
| Issue Date   | 2001-03                                                                       |
| Type         | Thesis or Dissertation                                                        |
| Text version | author                                                                        |
| URL          | <a href="http://hdl.handle.net/10119/906">http://hdl.handle.net/10119/906</a> |
| Rights       |                                                                               |
| Description  | Supervisor:奥村 学, 情報科学研究科, 博士                                                  |

# 論文間の参照情報の抽出と利用に関する研究

難波 英嗣

北陸先端科学技術大学院大学

2001 年 2 月

## 論文の内容の要旨

学術論文中には、当該論文と被参照論文との関係について記述されている箇所（参照箇所）がある。参照箇所から得られる情報を、本研究では参照情報と呼んでいる。本稿では、論文間の参照情報を自動的に抽出する手法を提案し、その応用例を示す。

引用分析において、これまで論文間の参照・被参照関係は様々な目的に利用されてきた。例えば、論文間の関連度の評価、研究領域の分析、論文誌や研究者の評価などが挙げられる。しかし、これらの多くはすべての参照を同等に扱っており、単純に参照の数を数えるだけでこのような分析を行うことに、引用分析の研究が始まった当初から批判があった。

そこで、本研究では論文間の参照情報に着目する。参照箇所からは、被参照論文の重要点や当該論文と被参照論文との相違点を明示する有用な情報が得られる。また、参照箇所を読めば参照の理由（参照タイプ）が分かる。こうした参照情報から、当該論文の関連論文の中での位置づけが明らかになるため、特定分野の研究動向の概要の把握に有用である。さらに、論文間の関連度の評価を含む多くの引用分析手法にも利用できると思われる。そこで、本研究では、まず論文間の参照情報の抽出を試みた。

参照情報の抽出は、参照箇所の抽出と参照タイプの決定という2つのステップに分けられる。本研究では参照箇所の抽出は、参照のある文と文間のつながりが強いと考えられる文を、参照の前後の文から抽出する処理であると考え、手がかり語を用いて参照箇所の自動抽出を行った。その結果、再現率 80%、精度 76%の抽出精度が得られた。また、抽出された参照箇所を手がかり語を用いて解析し、参照タイプを明らかにした。実験の結果、参照タイプ決定では 83%の解析精度が得られた。

参照情報の応用例の一つとして、サーベイ論文の作成支援を取り上げた。サーベイ論文には2つの処理 (1) 特定分野の論文の収集 (2) 論文間の類似点、相違点の抽出が必要であると考えられる。本稿では参照情報を用いることで、これらの処理を部分的に実現した。特定分野の論文の収集は、参照タイプを考慮して特定の参照・被参照関係を辿ることで可能になった。また、論文間の類似点・相違点は参照箇所中に記述されている。そこで、ユーザに論文間の参照関係を表すグラフ、グラフ中の個々の論文の概要、参照箇所を提示するシステムを構築した。このシステムを利用することで特定分野の論文が自動収集され、また収集された論文集合の論文間の相違点が明らかにされるため、参照情報がサーベイ論文作成支援に有用であることが示された。

また、参照情報の他の応用例として、関連論文の自動分類を取り上げた。書誌結合は引用分析の代表的な手法であるが、書誌結合ではすべての参照を等価に扱っているため、十分な分類精度が得られていない。そこで、本研究では参照情報を利用した関連論文の分類手法を提案し、実験により、提案手法と書誌結合を含む既存の手法を比較した。その結果、分類精度において提案手法が最も優れており、また計算速度においても、他の手法と比較して提案手法が十分高速であることが分かった。

本稿では、参照情報の応用例としてサーベイ論文の作成支援と関連論文の自動分類を取り上げたが、参照情報は引用分析研究で行われてきた研究領域の分析、論文誌や研究者の評価などにも応用できる。また、学術論文だけでなく、特許やウェブ文書といった他のジャンルのテキストへの適用なども考えられる。

キーワード： 引用分析、参照情報の抽出、複数論文の要約、文書分類