

特許間引用におけるオブソレッセンス分析

阪 明広

特許文献の引用の廃れの速さは本質的には発明の内容そのもの（内的属性）の影響で決まる。しかし、学術論文の引用分析において、引用は内的属性だけでなく、外的属性（特許文献で言えば、発明者数、ページ数など）の影響も受けることが知られている。このことは特許出願においても当てはまると考えられる。純粋に内的属性の影響を分析するためには、その前段階として、外的属性の影響を明らかにして、ベースラインを設定する必要がある。本研究はベースラインを設定するための準備として、基本的な知見（各属性の影響についての情報）を得ることを目的とする。

上記のことを明らかにするため、1993年に公開された特許を対象に、(1) 被引用半減期（被引用数の廃れの速さ）との相関の分析、(2) 被引用半減期が短い群と長い群の平均値の有意差検定、(3) Random Forests による、被引用半減期が短い群と長い群への自動分類という3つの分析方法を用いて、被引用半減期と外的属性を表す特徴量との間に関連があるかを調べた。分析には、NTCIR テストコレクションの公開特許公報の情報を用いた。調査した特徴量は、付与分類数、引用数、被引用数、ページ数、関数、表数、請求項数、発明者数、優先権主張数、優先権主張国数、出願してから公開されるまでにかかった日数（公開タイムラグ）である。

(1)に関しては、付与分類数、被引用数、引用数、公開タイムラグの4つの特徴量について、被引用半減期との有意な相関が観察された($p < 0.05$)。 (2)に関しては、被引用半減期が短い群と長い群の平均値に有意な差があると言える特徴量は、両群を分ける閾値を1~2とした場合には関数、3~4とした場合には付与分類数であった。 (3)に関しては、被引用半減期が短い群と長い群を分ける閾値を1から4まで変化させて調査したところ、被引用数、ページ数、関数の3つが、分類する上での重要な変数となった。

3つの分析を行い、色々な角度から、各特徴量の被引用半減期との関係を調査した結果、付与分類数および関数は、他の特徴量に比べると被引用半減期との関係が強いことが確認された。本研究では1993年に公開された特許を扱ったが、より長い期間について分析を行うことで、より有用な内的属性をより詳細に判断できるようにするためのベースラインの設定ができるだろう。

(指導教員 芳鐘冬樹)