

論文・特許のテキストデータを使った科学と技術の連関分析

文部科学省 科学技術・学術政策研究所
第2 調査研究グループ

要旨

本稿においては、1990年以降に出版された日本の著者による学術論文（約230万件）と日本特許庁に対する出願特許（約1200万件）のタイトル・要旨のテキストデータを用いて、科学（論文）と技術（特許）の相互連関関係について分析を行った。具体的には、それぞれの文献のタイトルと要約文を用いた分散表現ベクトルを作成し、コサイン類似度を用いた近傍文書の抽出し、論文の近傍特許数と特許の近傍論文数のトレンドや分野別特性を明らかにした。その結果、1990年代、2000年代、2010年代と時代が新しくなるにつれて、論文の近傍特許数は減少し、特許の近傍論文数は上昇するトレンドが見られた。これは、全体として、科学的なフロンティアの拡大が先に進み、技術的な進展が科学的な知見が多い分野をフォローする動きを表していると解釈できる。特許の非特許（論文）引用情報から、科学集約度の高い技術領域の抽出は行われてきているものの、本稿のアプローチによって、この科学→技術の関係に加えて、技術→科学（技術応用可能性が高い論文の学術領域の特定）の双方向の連関分析が可能となることを示した。

New indicator of science and technology inter-relationship by using text information of research articles and patents in Japan

2nd Policy-Oriented Research Group,
National Institute of Science and Technology Policy (NISTEP),
MEXT

ABSTRACT

In this study, the text information of academic papers (about 2.3 million) published by Japanese authors and patents filed with the Japan Patent Office (about 12 million) since 1990 are used for analyzing the inter-relationship between science and technology. Specifically, a distributed representation vector using the title and abstract of each document is created, then neighboring documents to each are extracted using cosine similarity. A time trend and sector specific linkage of science and technology are identified by using the count of neighbor patents (papers) for each paper (patent). It is found that the number of patents in the vicinity of papers decreased over time while the number of papers in the vicinity of patents increased. This can be interpreted as an advance the expansion of the scientific frontier by papers come first, then the technological progress (by patents) follows in the fields with substantial scientific knowledge already existed. The science intensity of technology has been measured by non-patent literature citation by patent. However, the citation information does not give the information of technology's impact on science. This study shows that our methodology enables both way interlinkage of science and technology.