

科学技術予測調査文書の視覚的分析

山田智子[†] 富樫 敦^{††} 藤井章博[‡]

Tomoko YAMADA Atsushi TOGASHI Akihiro FUJII

宮城大学大学院[†] 宮城大学事業構想学部^{††} 文部科学省科学技術政策研究所[‡]

1. はじめに

近年、ネットワーク分析技術やテキストマイニング技術など、計算機を用いた情報分析手法が、社会科学分野やマーケティング分野などでの分析に応用されている。同時に、技術経営や科学技術政策の策定において、経営者や専門家の主観や直感に頼るのではなく、客観的なデータを用いて戦略的な意思決定を実証的に行うことが重要視されており、その実現が期待されている。計算機を利用した文書情報の蓄積・検索技術の高度化を背景に、洗練された文書情報の分析が可能になり、科学技術政策形成の分野においても、分野の動向に関する知見の提供、研究開発活動の多様性に対応した科学技術に関する知見の提供が可能になっている。

政府が行う大学等への公的な研究開発投資は、日本全体の約 2 割であり、国全体の研究開発を牽引しているといえる。そのための重要な課題は、科学技術の多岐にわたる分野の動向を多角的に分析し、科学的に検証し、効果的な投資に結びつけることである。

従来、科学技術政策形成において、各科学技術専門家によって検討され科学技術動向にもとづいて、ステークホルダー（政策意思決定者）による政策上の意思決定が実施されている。科学技術予測調査はエキスパートの知識から科学技術政策における意思決定に有用な知見を研究・開発投資の効果的な割り当てのためのステークホルダーへの知識提供という面も持っている。その一方、科学技術分野の細分化、学際化により専門分野外の現状や動向を把握することは非常に困難であるという現状がある。

2. 調査資料に基づく技術動向分析法

2.1 技術動向分析のプロセス

科学技術政策に関する意思決定を行うためには、科学技術の個別分野において、既知情報に基づいた技術展開を予測しなければならない。近年、科学技術知識の高度化と複雑化にともな

い、技術展開予測の機械的な支援が望まれる。そこで、技術発展の動向などをモデル化するために、蓄積された文書群が、どのように互いに関連付けられるかという点を考察する。

文書として蓄積されるものは技術発展に関する記述である。ここでは、文書に出現する単語の頻度に基づいて、文書を互いに関係付けることにより、技術発展を把握することを試みた。本研究では、分析対象文書に関して、各分野を特徴付ける語彙（キーワード）を機械的に抽出することにより、文書解析基盤を構築した。次に、本研究で提案する「シソーラス分析」と「ネットワーク分析」を組み合わせた手法により、対象とする科学技術予測調査文書を計量的に分析した。

最初のキーワード群の抽出ステップではオントロジーの考え方にに基づき、適切な語彙抽出した。計量分析のステップでは、web 上で公開されているテキストマイニングツールを利用し、キーワードの頻度ベクトルによるベクトル空間法に基づき文書間の関係を定量的に分析した。分析結果を可視化するため、ネットワーク分析手法を用いて、文書間の関係を図的に明示した。

2.2 領域依存のタクソノミー構築

本研究では、科学技術の個別分野の特徴を、キーワード（語彙）群に基づいて分析することとする。ここで、科学技術分野を特徴付けるようなキーワード群が適切に抽出されなければならないが、一般の形態素解析だけでは、適切な複合名詞を機械的に抽出するには限界がある。従って、科学技術を特徴付けるキーワード（語彙）群を自動生成することは難しい。

以上から、分析の精度を高めるために、個別の科学技術分野に応じたキーワード（語彙）を作成することを目的として、領域オントロジー構築[8]の考え方を取り入れ、科学技術各分野の適切な語彙抽出を行い、科学技術分野のタクソノミーを構築する。ここで、タクソノミーとは概念階層定義のことであり、オントロジーとはタクソノミーに概念間系が追加されたものである。分析対象分野のキーワード抽出の概要を、図 1 に示す。

「Visual Analysis of Technology Prediction Investigation Document」

[†] Graduate School, Miyagi University

^{††} School of Project Design, Miyagi University

[‡] National Institute of Science and Technology Policy

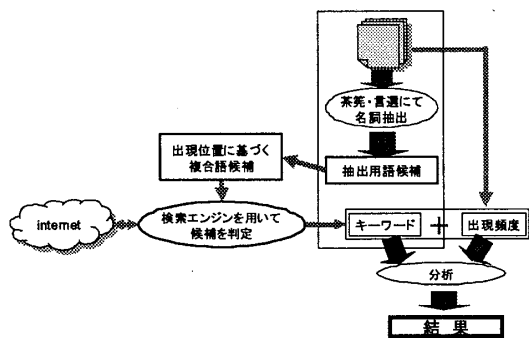


図1, 分析対象分野のキーワード抽出概念図

具体的には、次の手順で、タクソノミーを抽出した。

- ① 分析対象全文書に対して「Windows 用テキストマイニングツール termmi」^[4]と専門用語（キーワード）自動抽出用 Perl モジュール「TermExtract」^[5]を使い、専門用語（複合名詞）を抽出した。
- ② 抽出された専門用語からキーワードを選別し、各文書に対して、出現頻度を出した。
- ③ キーワードの出現頻度から、文書の特徴を分析した。

ただし、語彙間の構造は、今回は、考慮しないものとした。

3. 科学技術予測調査資料を対象とした分析

本節では、2. の手法により科学技術予測調査資料を分析することで、各科学技術分野の動向を把握することが可能となることを確かめる。具体的には、分析として、図2で示したプロセスにより分析を行うが、特に、キーワード生成について、領域依存のタクソノミーの有効性を考察するため、2種類のキーワード群で分析を実施した。

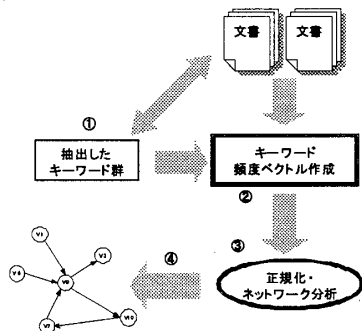


図2, 調査資料に基づく技術動向分析プロセス図

分析対象文書から出現頻度上位でのキーワード群と専門用語（キーワード）自動抽出用 Perl モジュール「TermExtract」を使い専門用語抽出しキーワード群を作成した。

結果、2つのネットワーク図が生成した。出現頻度上位のキーワード群、専門用語のキーワ

ード群にし、分析をした。

分析の結果では、関連度が低いと思われていた文書間の関係が、キーワード群を専門用語にしたことで、図5から明らかなように、文書間の関係が鮮明になったことがわかる。科学技術の当該分野の特徴をキーワード（語彙）で表し、分析対象とした。ここでは、オントロジー構築の考え方を採用し、科学技術各分野を表すキーワードが抽出可能となり、分析の質的向上が図られたと考えられる。

4. おわりに

本稿では、科学技術予測調査文書を事例として、文書間の関係を定量的に分析し、文書の特徴を抽出した上で相互関係を把握し、特徴を理解することを目的に分析を実施した。分野を特徴づける語彙に関して「タクソノミー構築」を実施し、対象となる文書の特徴と相互関係を検証した。分析の対象として、文部科学省科学技術政策研究所が実施した「科学技術の中長期発展に係る俯瞰的予測調査」で蓄積され、一般に公開されている資料を利用した。分析は、科学技術政策の観点で特定の科学技術分野の変化を視覚的に把握する目的で行った。近年注目されているネットワーク分析の手法が、こうした文書間の関係、文書の分析にも有効であることがわかった。分析では、基本的な考え方に従い、インターネット上のサービスを援用したもの、主として手作業でキーワード群を生成したが、インタラクティブなツールなどを利用することにより、さらに、分析の効率が高まるものと期待できる。

謝辞

本研究は、(財)新技術振興渡辺記念会科学技術研究助成、笹川科学研究助成による支援を受けている。

参考文献

- [1] 「科学技術の中長期発展に係る俯瞰的予測調査 - 注目科学技術領域の発展シナリオ調査 -」 NISTEP REPORT No.96 2005.5 科学技術政策研究所
- [2] UCINET <http://www.analytictech.com/>
- [3] Windows 用テキストマイニングツール termmi <http://gensen.dl.itc.u-tokyo.ac.jp/termmi.html>
- [4] 専門用語（キーワード）自動抽出用 Perl モジュール「TermExtract」 <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>
- [5] 安田雪「ネットワーク分析」2002 新曜社
- [6] 安田雪「実践ネットワーク分析」2003 新曜社
- [7] Takeshi Morita, Naoki Fukuta, Noriaki Izumi, Takahira Yamaguchi, "DODDLE-OWL: A Domain Ontology Construction Tool with OWL", Proceedings of the 1st Asian Semantic Web Conference Lecture Notes in Computer Science Vol.4185 pp.537-551, (2006.9)