

科学技術政策を対象とした文書の知識型計量分析手法の研究

宮城大学事業構想学部 富樫 敦

1. 研究課題全体について

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、社会科学研究分野やマーケティング分野などの研究分野で応用されている。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。

そこで、企業戦略策定などでは客観的なデータに基づき検証可能な意思決定を行うことが重要であるとの観点から、本研究では、シソーラス分析とネットワーク分析を組み合わせた知識型テキストマイニングにより文書を計量的に分析する手法を提案し、科学技術政策立案分野での適合性の検証を行う。

ここでは、専門家に対するヒアリング調査の結果と機械的な情報分析をリンクさせるということを目的として、文部科学省科学技術政策研究所が実施した技術予測調査「科学技術の中長期的発展に係る俯瞰的予測調査」を事例として取り上げた。そして、「注目科学技術領域の発展シナリオ調査」で蓄積された科学技術予測文書を対象とし、各科学技術分野の特徴分析を行った。

特に、情報分析では、専門用語辞書の品質が、結果に大きく依存するため、適切な専門用語の抽出法を検討した。具体的には、インターネット上の検索エンジンを補助的に使いながら、重要な専門用語であるかどうかの判断を支援し、専門用語辞書を半自動的に構築することにより、分野ごとの専門性や、分野横断性を、分析結果に正確に反映する手法を提案した。

本研究の評価として、専門用語抽出の有効性を検証するために、二種類のキーワード群を使い、比較・検証を行った。結果として、文書間の関連性が低いと思われていた関係が、適切な専門用語辞書を用いることで、文書間の関係として正確に可視化できることがわかった。さらに、一部の分野に関して、専門家による定性的評価を実施し、本研究の分析手法の評価・検証を行った。

2. 研究概要

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、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専門用語（語彙、あるいはキーワード）を半自動的に抽出する方法論を提案する。さらにこの方法論に基づき、抽出された専門用語群を次元とする頻度ベクトルを各文書に対応させ、内積等により、文書間の関係をより適切に分析するための基盤を構築する。

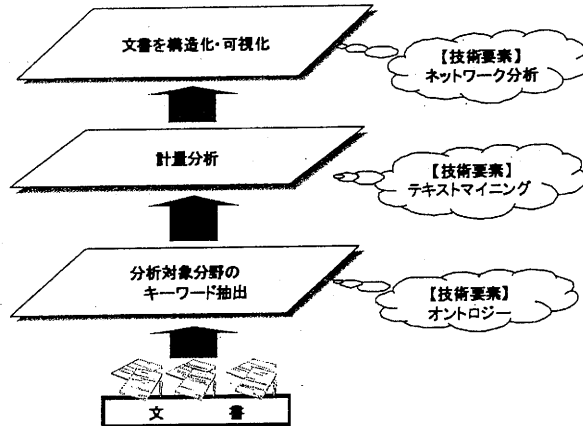


図 1 分析プロセス概念図

本研究では、一般文書から専門用語を抽出するためのアルゴリズムを考案し、これに基づいて抽出された専門用語を用いて文書の分析を行う。本研究で行う文書分析法はベクトル空間法の一つであり、その結果はネットワーク図等を用いて可視化する。本手法（専門用語抽出法と文書分析手法）の妥当性を検証するため、科学技術分野の動向や今後のシナリオを述べた技術文書を対象にした実験を行う。

本研究で提案実証する分析手法では、各科学技術分野の特徴を語彙（当該分野を代表する専門用語）で表し、科学技術文書の分析で用いている。本研究では、分析の精度を質的に高めるため、語彙の作成に情報工学や人工知能分野の新技术「オントロジー技術」の概念を取り入れ、科学技術各分野の語彙の体系化を行い、分析の信頼性を高める。

3. 分析対象分野語彙抽出

本研究で提案実証する分析手法では、各科学技術分野の特徴を語彙（当該分野を代表す

る専門用語)で表し、科学技術文書の分析で用いている。本研究では、分析の精度を質的に高めるため、語彙の作成に情報工学や人工知能分野の新技术「オントロジー技術」の概念を取り入れ、科学技術各分野の語彙の体系化を行い、分析の信頼性を高める。

4. 科学技術予測調査資料の相関分析

発展シナリオ調査から得られた、48テーマ85文書の発展シナリオを定量分析し、発展シナリオ間の相互関係を分析する。

科学技術を特徴づける語彙を抽出し、発展シナリオ文書毎に抽出した語彙の頻度分析を行う。頻度分析の結果から、85文書の発展シナリオそれぞれの特徴を定量的に計測し「頻度ベクトル」で表す。「頻度ベクトル」を用い、ネットワーク分析で発展シナリオ間の関係を可視化技術を使いネットワーク図で視覚化し、結果を提示し、動向を分析する。

また、本分析では、キーワード抽出に関して、領域依存のタクソノミーの有効性を検証するため、2種類のキーワード群を使い分析結果の検証も行った。

5. デルファイ調査結果と発展シナリオ調査結果の対応関係の分析

デルファイ調査結果と同調査を補完する発展シナリオ調査結果の相関を計量分析する。第8回のデルファイ調査では13分野を設定し、「分野-領域-予測課題」の階層構造でその分野を特徴づけている。

本研究では、「領域」をキーワードの固まりと捉え、分析の対象とする。本研究では、デルファイ調査結果からキーワード群を抽出し、発展シナリオ調査にひそむ2種類の文書間の関係性を分析する。

6. 評価・検証

本研究で提案した分析手法の評価を専門家に協力をしていただき、定性的評価を実施した。評価方法としては、対象文書を内容を理解していただいた上で、専門家の知見で主観的に文書間の関係をネットワーク図として作成していただき、本研究で分析した分析結果と比較・検証を実施した。

また、デルファイ調査で設定された13分野に対して、専門用語の抽出を行い、抽出された専門用語の品質評価を行った。

評価対象としては、13分野に対して評価を行うのが難しいため、13分野の中から、「情報・通信」、「保健・医療・福祉」に対して、専門家に定性的評価を実施した。