JICSTファイル・特許DBを用いた動向分析

—データマイニング手法を用いた技術連関分析—

中村達生¹

著者抄録 : データマイニング手法を用いて、論文と特許の内容的概念的類似性からつながりを示す方法を提示し、人工知能分野、液晶分野、バイオテクノロジー分野の論文について、全特許文献との技術連関分析を行うとともに、本手法の有効性を示した。データマイニング手法の俯瞰（ふかん）的な分析力を生かして、論文と類似特許の件数について年別推移傾向を表し、論文の隆盛（減少）傾向と特許の傾向が、非常に似ている技術分野が存在することを明らかにした。また、類似特許のIPC分類に着目して、各技術に対するコア技術と周辺技術を同定した。

キーワード : データマイニング, 技術連関分析, 自然言語処理, 特許分析

Using JICST file and Patent DB on a technology trend analysis

—Using data-mining on inter-technology analysis—

NAKAMURA Tatsuo¹

Author Abstract: This paper demonstrates the approach and the availability of data mining for inter-technology analysis. AI, LC and biotech were chosen for this trend analysis between research paper and patents. Depending on the merit of data mining that make it possible to search from all data, the trend analysis of papers and patents numbers was success and some conformity was found. And a classification to core and other technology was done using IPC code counting.

Key words: data mining, inter-technology analysis, natural language processing, patent analysis

原稿受理 (2003-03-20)

(情報管理 46 (2): 97-106)

1. 研究の目的

科学技術立国を標榜して久しいが、わが国産業の国際競争力を強化し、経済を活性化していくために、研究活動や創造活動の成果を知的財産として戦略的に保護・活用していくことが今ほど必要とされていることはなかった。平成14年11月27日に成立した科学技術基本法の中でも、経済社会において知的財産が積極的に活用されることをうたっている。特に、大学や研究機関で創出された基礎技術が、産業、製品へとつながる様がどのように連関しているのかを分析し、明らかにすることは、今後の技術政策を行ううえでも非常に重要である。

そこで本研究では、データマイニング（概念検索）手法を用いて、論文と特許の類似性に基づく技術連関分析を行う方法を提示し、その手法の有

効性と適用の可能性を示すことを目的とする。検討にはバイオ分野とIT分野を対象として、サイエンス（論文）と産業技術（特許）の時系列的な推移の傾向分析と、技術分野的な広がりの実施した。なお本報は、文部科学省科学技術振興調整費により、平成11～12年度にかけて実施された調査の一部を核として報告するものである。

2. 従来の研究

2.1. 従来技術連関・波及効果分析手法

従来技術連関に関する研究は、テクノロジー・トランスファーに関する事例研究を挙げることができる。これらはスピノフ、スピノン、スピルオーバーによる技術移転の事例研究であり、特に米国における軍事技術の民生技術への転用事例

¹(株)三菱総合研究所 (〒100-8141 東京都千代田区大手町2-3-6) Tel. 03(3277)0502 E-mail: tatsuo@mri.co.jp

¹Mitsubishi Research Institute, Inc. (3-6, Otemachi 2-chome, Chiyoda-ku, Tokyo, 100-8141)

の研究が多い。

定量的な分析では、サイティングインデックスを使った分析手法がある。ISI社の提供する引用文献データベースを用いて、論文の引用件数の大小からつながりの強さを示そうというものである。最近では、引用度の大きい論文ほど重要な論文であるという仮定に基づき、研究開発プロジェクトの評価の1項目として取り上げる動きも出てきている。しかし、論文のレファレンスに記載される引用論文リストは、分野により差異が大きいことや、評価に直結するということからグループ間で相互引用をするなど恣意（しい）性があることも指摘されている。

定量分析の別の方法として、論文に付与されたキーワード群の組み合わせの共起確率を基に、つながりを推計する方法もある^{1,2)}。論文には主題となるキーワードのほかに、副主題のキーワードが付与されており、その組み合わせの頻度を集計することで、分野間のつながりを知ることができる。しかし、この手法はキーワードが正しく付与されていることが大前提であり、それは情報サービス提供会社の処理精度に依存してしまう。

上記の研究は論文間の技術的なつながりが分析の対象であるが、米国では特許中に書かれた論文の件数を集計した特許と論文のサイエンスリンクエージ（Science Linkage）という指標に基づく分析がなされている。米国特許では特許中の参考文献リストを記載することが制度的に規定されており、代表的な研究者としてCHI ResearchのDr. Narinを挙げることができる。Dr. Narinの研究によれば、米国に出願された日本企業と欧米企業の特許を用いて比較すると、日本企業の出願した特許は引用した論文の数が極めて少なく、すなわちサイエンスリンクエージが突出して低いという結果が得られている。特許のサイティング情報を用いると、技術連関のみならず、背景にある研究開発上の連携体制なども含めた国際比較分析が可能となるが、日本国特許には参考文献の記載規定がないために、このままでは網羅的な分析をすることができない。

2.2. 俯瞰的な技術連関分析を可能とするデータマイニング（概念検索）手法

全技術分野間の網羅的なつながりを分析するには、分野ごとの専門的な知識が必要とされるが、技術的なつながりを概念的な類似性に基づくものと定義すれば、データマイニング手法を使うことで、機械的に抽出することができる。

本研究のデータマイニング手法は概念検索とも呼ばれており、全文検索である。したがって、キーワードの付与漏れによる抽出誤差は基本的に発生しない。論文の抄録を入力文として、特許データベースを検索すれば、入力した論文と類似した特許を抽出し、その類似性に応じた指標を用いて序列化することができる。直接関係のある特許のみならず、間接的に関係のある特許も含めて抽出され、さらに統計処理を施して、技術間のつながりを定量的に示すことができる。

3. 分析方法

3.1. データマイニング手法の基本的な考え方

分析に用いるデータマイニング手法は、あらかじめデータベース中の各文章をベクトルで表現しておき、入力した文章（論文）のベクトルと方向が近いものほど類似しているとして検索を行っている。検索時においては、入力文とデータベース（以下、DB）中の文書の内積を計算するだけとなり、大量の文書検索においても、従来のテキスト検索と比較して高速である。

文章をベクトル化するには、形態素解析と呼ばれる方法を用いて複数の単語に分割し、各単語の重要性は、文章とDB中に現れる頻度から決定する。ベクトルの近さは類似度と呼ばれる指標で表され、この数値の大きい文章ほど概念的に近いとされている。

(1) 形態素解析

形態素解析とは、文章を辞書に登録されている語へと分解する処理を意味し、大きく分けて 1) 語の切り出し 2) 接辞処理の2つの段階から成る。

(2) 単語の重み付けの方法

文章中の各単語の重みは、文章に現れる頻度と文献中にその単語の含まれる文書の偏在率を用いて算定する。すなわち、文献の中に繰り返し出現する語であり（TF）、かつ、その語を含む文献が少ない（IDF）ほど、ウエートが大きくなる（式(1)参照）。文書中の頻度（TF）のみの考慮では、機能語や一般語の重要度も高めることになり、有意な検索ができなくなる。

(3) 類似度

検索質問（入力文）と各文献の類似度の算定は

$$W_{ij} = TF_{ij} \times IDF_j \cdots \cdots \text{式(1)}$$

TF_{ij} : Term Frequencyの略。文献iの中に出現する語jの頻度

IDF_j : 逆文献頻度(Inverse Document Frequency)の略。

語jを含む文献数の逆数

内積の考え方を基本に用いている(図1)。ベクトルは文献中の各単語を独立した次元として、これらの重みを要素としている。類似度の計算式には目的に応じたいくつかの方法があり、また類似度はRSV(Retrieval Status Value: 検索状態値)のひとつであり、RSVに関する既存研究は、主に自動索引作成を目的として実施されてきた³⁾。

3.2. 技術関連分析の方法

本研究のデータマイニング手法を用いた技術関連分析では、ある特定の分野についてサイエンス(論文)と産業技術(特許)の時系列的な推移傾向と、技術分野の変化に関する分析を実施した。

論文をサイエンスとしてとらえて、同一分野の論文の抄録の内容と類似する産業技術の特許文献の中から抽出する。次に、抽出された特許文献の類似度や出現件数を出願年別に整理し、入力に用いた論文群の発行年別件数推移と比較分析を行う(図2)。さらに、抽出された類似特許に記載されたIPC分類に着目し、同じく出現頻度や類似度から、コア技術や周辺技術を抽出するとともに、関連するIPC分類構成の変化を解析する。IPC分類の数が増加している場合は、波及分野が拡大していることを示し、逆に、関連するIPC分類の数は限られて

図1 ベクトル空間モデルのイメージ（三次元の例）

いるが、出現頻度が増加している場合は、技術が深度化しているものと推察できる。

3.3. 検討対象分野と分析に用いた文献データ

(1) 分析対象分野

分析の対象分野は、過去の研究経緯が既存の調査結果から比較的把握しやすい分野であることが、本手法の検証という点で望ましく、情報分野のソフト技術である「人工知能技術」、情報分野のハード技術分野として「液晶技術」を選択した。これに、ゲノムに代表されるように最近の研究開発競争が激しく、論文件数が急増しているバイオテクノロジー分野を加えて全3分野を対象とした。

(2) 対象文献データ

検索入力文に、分野別論文の書誌情報(タイトル+キーワード+抄録)とし、データマイニング分析システムを用いて、特許文献データベースから類似する特許文献を抽出する。論文データは、JICSTファイルを用いて科学技術分類とシソーラスを用いて、人工知能、液晶、バイオテクノロジーに関する論文の抄録を含めた書誌情報を抽出した。論文はJICSTファイルの分析に供することができる全期間(1980~2000年)を対象とした。一方、特許データは、電子ファイル化された全特許(1993~2000年9月登録分)を対象とした。出願年ベースでは、1974~1998年の期間を収録している。

3.4. データマイニングによる類似特許の検索例

データマイニング手法により論文に類似した特許文献が抽出されることを確認するために、人工知能分野から2つの論文を選定し、それぞれに対して類似度が最も高くなる特許文献の概要を整理して、その類似性を比較検証した。検証に用いた2編の概要は表1のとおりである。

図2 論文データと特許データベースの入出力関係

論文Aに対する類似特許文献は、「パターン・データ処理装置」(特願平4-186878)であった。論文Aの抄録と特許文献の明細書の中から特許の特徴を端的に表す項を比較すると、表現方法は異なるが、内容的に非常に類似していることがわかる(表2の下線部分)。

また、論文Bに対する類似特許文献は、「信号処

理装置」(特願昭63-23285)である。当該論文と特許文献はともに、階層型(特許では多層型)ニューラルネットワークにおける逆伝搬法(特許ではバックプロパゲーションと表現)等の学習則を研究(開発)対象としていることが共通であり、類似度が高く計測された主な要因であると考えられる(表3,表4)。この特許文献が当該技術を信号処理装置

表1 サンプルとして用いた論文の書誌情報

論文	書誌情報
論文A	松山隆司(京大工),“画像理解におけるトップダウン・ボトムアップ解析の統合”,電子通信学会技術研究報告, VOL. 84, NO. 248 PAGE. 73-84(AL84-47) 1984
論文B	横井博一(九州工大 工); 木本智幸(大分工高専),“介在素子を付加した階層型ニューラルネットワーク”,バイオメディカル・ファジィ・システム学会誌, VOL. 1, NO. 1 PAGE. 87-97 1999

表2 論文Aの抄録抜粋と類似特許の明細書抜粋

◎論文(人工知能論文A)の主要ポイント SIGMAでは、部分情報の蓄積・統合によってボトムアップ、トップダウンの二つの解析過程を統合し、柔軟な制御方式を実現
◎特願平4-186878の主要ポイント 【発明の詳細な説明】よりを抜粋 【0021】本発明は、上記事情に鑑みてなされたもので、トップダウン処理とボトムアップ処理を有機的に結び付けて精度の高いパターン認識を実現できるパターン・データ処理装置を提供することを目的とする。

表3 論文B「介在素子を付加した階層型ニューラルネットワーク」の抄録とキーワード

<抄録> 階層型ニューラルネットワークにおいて、その中間層の素子数を増やせば表現能力が向上する。しかし、誤差逆伝搬学習則を用いる場合、ローカルミニマムの存在のため、素子数の増加は必ずしも学習能力の向上にはつながらない。本論文では、小脳皮質や大脳皮質のネットワーク構造をヒントに、ネットワークの層間に介在素子を有する階層型ニューラルネットワークを提案し、その一般的な学習則を導いた。介在素子として、特徴検出素子とカテゴリ化素子の2種類を提案した。次に、3層ニューラルネットワークに対して、介在素子無しの場合、特徴検出素子を付加した場合、カテゴリ化素子を付加した場合の3通りの計算機シミュレーションを行った。学習能力を調べた結果、介在素子を付加した方が、中間層の素子数を増やすよりも高い学習能力が得られた。また、特徴検出素子と比べて、カテゴリ化素子を用いた方が少ない素子数で同等の学習能力が得られた。
<技術キーワード> 神経回路網モデル; 学習曲線; 計算機シミュレーション; 階層構造; 特徴抽出; 逆伝搬法; 多層構造; カテゴリ

表4 「信号処理装置」(特願昭63-23285)の抜粋内容

【発明の詳細な説明】 A 産業上の利用分野 本発明は、学習機能を有する信号処理装置に関し、特に、所謂ニューラルネットワーク(Neural Network: 神経回路網)によるバックプロパゲーション(Back propagation: 逆伝播)学習則に従った学習機能を有する信号処理装置に関する。 B 発明の概要 本発明は、ニューラルネットワークによるバックプロパゲーション学習則に従った学習機能を有する信号処理装置において、ニューロンに対応する各ユニットによるリカレント回路網にて入力信号に対する信号処理を行うことにより、上記リカレント回路網の結合状態の学習にて、時間軸上で変動する音声信号等のシーケンシャルな時系列パターンの特徴を抽出できるようにしたものである。 C 従来の技術 ニューラルネットワークの学習アルゴリズムであるバックプロパゲーション学習則『「Parallel Distributed Processing」Vol.1 The MIT Press 1986や日経エレクトロニクス1987年8月10日号, No. 427, pp115-123等参照』は、第4図に示すように、入力層(31)と出力層(33)の間に中間層(32)を有する多層構造のニューラルネットワークに適用され、高速画像処理やパターン認識等の各種の信号処理への応用が試みられている。 <途中省略>
--

に適用することを主題としているのに対し、論文のほうでは、当該技術による学習能力の比較を主題としていることに相違点があるが、既知である当該技術から研究開発対象が広がった可能性を示唆しているとも見ることができる。

以上の検証結果から、データマイニング分析では、論文の概念を用いて、主要骨子がほぼ一致する特許の抽出に成功していると言える。

4. 論文件数と類似特許の件数推移分析

論文の件数推移と、類似する特許文献の件数推移の分析を、対象とした3つの分野の要素技術について適用した。論文が基礎技術の成果、特許文献をその応用である産業技術の成果とすると、両推移の共通性から対象分野の研究開発動向を推測することができる。

分析に供する各分野の要素技術は、技術動向、社会潮流等をかながみてそれぞれ複数のテーマ(表5)を選定したが、ここでは誌面の都合上、人工知能分野からは「第5世代コンピュータ」、液晶分野からは有機EL、バイオテクノロジーからは遺伝子組み換え技術について分析の結果を紹介する。

4.1. 第5世代コンピュータの論文と特許の件数推移

(1) 第5世代コンピュータプロジェクトの概観

第5世代コンピュータとは、通産省(現経済産業省)主導の下で1982年4月から10年間実施する計画で始められた「第5世代コンピュータプロジェクト」の成果目標である。1980年に商業化された「第4世代コンピュータ」が大規模周期回路(LSI)を用いていたのに対して、「第5世代コンピュータ」は単に使用する半導体素子を高度化するだけではなくコンピュータの仕組みそのものを変えようとした。その特徴が並列処理能力、知識のデータベース化、自然言語(日本語)での入出力機能を備えることであった⁴⁾。第5世代コンピュータプロジェクトは、ICOT^{注1)}という独立した組織が政府から委託研究を受ける形で行われ、当初の成果は国のものになった。1991年から、成果は参加企業と国の共有となった。

第5世代コンピュータプロジェクトは企業がまったく出資しなかったため、予算が計画の半分になってしまい、途中から研究テーマは「並列推論マシン」(Parallel Inference Machines PIMs)の研究開発に絞り込まれた。

表5 分野別検討対象技術キーワード

技術分野	対象技術キーワード
人工知能	① 自然言語処理
	② エキスパートシステム
	③ マルチエージェントシステム
	④ データマイニング
	⑤ 知能ロボット
	⑥ 第5世代コンピュータ
液晶	① スメチック層
	② ドットマトリクス
	③ 有機EL
	④ Active matrix
バイオテクノロジー	① 遺伝子操作
	② ヒト成長ホルモン
	③ バイオセンサー
	④ 遺伝子組換え技術
	⑤ ゲノム*解読
	⑥ エンドルフィン
	⑦ モノクローナル抗体
	⑧ ヒトインターフェロン

当初の予定より1年延長して11年になったが、1990年代のメインフレームコンピュータをつくるという当初の目的は達成していない。

「逐次推論マシン」は三菱電機と沖電気によって製品化されたが、既存のワークステーションに価格性能比で勝てず、市場シェアを伸ばせなかった。どの参加企業もプロジェクトで開発した「並列推論マシン」を商業化する予定を持ちえなかった。

第5世代コンピュータプロジェクト開始後に、コンピュータ産業ではパソコンやワークステーションが驚異的な進歩を遂げ、オープンシステムの浸透とダウンサイズ化が進み、メインフレームの重要性も低下せざるをえなかった。

(2) 論文と特許の件数推移の分析結果

第5世代コンピュータプロジェクトに関する論文件数の推移は1980年代の間、プロジェクトが実施されていた期間中は一定の盛り上がりを見せているが、プロジェクトの終了とともに1992年ごろから急激に減少している。特許類似度推移もこれとほぼ同じ推移をたどっており、研究開発が同時期に終えんしていることを示している(図3)。

4.2. 有機ELの論文と特許の件数推移

(1) 有機EL (Electro Luminescence)

有機ELは自己発光の素子であり、それを用いたディスプレイは、LCDのようにバックライトを必

注1) 財団法人新世代コンピュータ技術開発機構 (Institute for New Generation Computer Technology)

図3 第5世代コンピュータの論文と類似特許の件数推移

要としない。また、視野角の制限もなく、応答速度もLCDより優れている。低消費電力で高輝度、薄型化が図れるため、デジタルカメラや携帯端末機器 (PDA) 等に応用すると、高品位表示で、なおかつ駆動時間の延長や機器の薄型化が実現できると期待されている。中・小型の分野では、今後LCDに代わる次世代ディスプレイとして注目されている。

(2) 有機ELの論文と類似特許の件数推移

有機ELに関する論文数は、1996年以降になって急激に増加してきている。主な著者は基本的に大学関係者が多いが、三洋電機、出光興産、日立、沖電気、NECなどの企業研究者による発表も目立つ。一方、類似特許は1993年に集中して高い類似度の特許が住友電気工業(株)をはじめ各社から多数出願されている。キヤノン(株)は有機ELの名称はないが、有機ELと同等の機能に関する特許も類似特許として抽出されている。

論文と特許の件数はほとんど同形で推移し、1995年以降に急増している。グラフでは1999年に件数が激減しているように見えるが、分析当時において1999年のデータが未整備であったためであり、実際には1999年も増加中と考えられる。また、特許の件数推移にのみ1992年にもピークが発生しているが、特許法の改正によるかけこみ出願と関係があるものと推察される (図4)。

4.3. 遺伝子組み換え技術の論文と特許の件数推移

遺伝子組み換え技術の論文件数は1995年までゆるやかな増加傾向にある。一方、類似特許文献の件数推移は1988年をピークにして激減している。その後の増減傾向は論文とほぼ同じであるが、減少傾向をたどっている (図5)。

4.4. 論文と類似特許文献の件数推移の分析結果

液晶、人工知能、バイオテクノロジー分野の個別技術に関する論文と、その論文の概要と類似した特許の件数の推移を比較した結果、論文の隆盛 (減少) 傾向と特許の傾向が非常に似ている技術分野の存在することが明らかとなった。これはエキスパートシステム、第5世代コンピュータ、ドットマトリクス、有機EL、ヒト成長ホルモン、モノクローナル抗体、ヒトインターフェロンであった。論文を著し、学際的研究が盛り上がると、ほぼ時を同じくして、特許も類似概念の出願件数が増加していることを示している。この中には、エキスパートシステムや第5世代コンピュータのように、過去に研究のピークを有した技術分野もあれば、有機ELのように、1990年代後半から急激に増加傾向を示すものも含まれている。

図4 有機ELの論文と類似特許の件数推移

図5 遺伝子組換え技術の論文と類似特許の件数推移

5. 論文から特許への技術分野の広がり分析

分析対象の3分野に類似した特許文献に付与され

ているIPC分類を、出願年別に件数を集計し、そのコアとなる技術と周辺技術を明らかにした。出現件数の順に並べたとき、出願年によらず有意に大きい値をとるIPC分類をコア技術とし、それ以外に

表6 エキスパートシステムにおける類似IPC分類の年別推移

IPC	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
G06F		2	1	6	4	1	8	7	13	18	11
G05B	1	1	3	3	1	3		2	5	5	2
G21C				1		1		2	3		1
H01L	1	3						1			2
A01N							8				
H02J	1		2					1			
A61B		1					1			1	
G10L											
G06T				1					1	1	
C07D							2				
G01N							1		1		2
G21F									2		
B23Q			2	1				1			
H01M	1		1	2							
H04N	2									1	
C12P			2				2				
C12N			2								2
E04G							2			1	
C09K		1									
G09B								1		1	
G01D									1		
G08B								1			1
B01J							1				
C12R			1				1				1
H04L								1			
E01C											
C21B								1			
G05D			1								1
B66B							1	1			
E02D		1									
A61K							1				
G11B	3										
IPC	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	合計
G06F	2	12	13	4	3	3	6	4	3		121
G05B		2	3		1				3		35
G21C	1	2	2						3		16
H01L		2	4								13
A01N											8
H02J		3									7
A61B							2	1			6
G10L					1		1	1	3		6
G06T		2							1		6
C07D			2			2					6
G01N		2									6
G21F		1	1						1		5
B23Q			1								5
H01M	1										5
H04N			1						1		5
C12P									1		5
C12N									1		5
E04G					1	1					4
C09K			3								4
G09B		1		1							4
G01D			1						1		3
G08B		1									3
B01J								1	1		3
C12R											3
H04L			2								3
E01C	1		1						1		3
C21B	1								1		3
G05D									1		3
B66B			1								3</

表7 エンドルフィンにおける類似IPC分類の年別推移

IPC	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
C12P	4	4	8	11	16	33	55	37	44	27	40	40
C12N	3	5	6	6	16	29	43	54	45	26	38	39
C07K		6	3	6	8	24	35	21	39	26	23	46
A61K	1	1	4	4	12	22	27	19	23	29	30	36
G01N	1		2	7	8	16	28	11	37	19	24	34
C12R	1	2	2	5	10	23	34	35	28	14	24	20
C07D				1						2	4	4
C12Q	1				3	2	1	4	9	1	1	4
A61P					1	1				1	2	4
C07H					3	2	3	1	3		2	1
B01J								1				
C07C												1
C07F										1	1	
C07M											1	
C07J									1			
A61M						1					1	
A01K										1		
C08B									1			1
C11D									1			
A01N												
A23L										1		
A23K										1		

IPC	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	合計
C12P	23	3	6	14	14	2	1		1		383
C12N	14	4	10	19	11	7	2		1		378
C07K	7	15	6	12	7	5	3		1		293
A61K	11	17	13	8	15	5		3	2		282
G01N	13	11	9	13	9	2	3				247
C12R	13		3	5	10	3					232
C07D	3	8	3	2	4	6	1	1	2		41
C12Q	2	2		2	2	1					35
A61P	2	3		2	1	4		1	2		24
C07H	1		1	1							18
B01J		1	4		2						8
C07C					3			1			6
C07F		1									3
C07M			2		1						3
C07J											2
A61M			1								2
A01K		1									2
C08B											2
C11D					1						2
A01N				1	1						2
A23L						1					2
A23K						1					2

出現するIPC分類を周辺技術としてグループ化を行った。分析に際して整理した表（表6、表7）の縦軸は、類似する特許に付与されているIPC分類のセクションレベルであり、横軸が類似特許の出願年を表している。表中の数字は、類似している特許の件数を表しており、この数値が大きいほど、個別技術の論文と特許の該当分野と概念的な類似連関が強かったと言える。人工知能分野のうち、エキスパートシステムに関する技術では、コア技術がG06F（電氣的デジタルデータ処理）であり、これはエキスパートシステム特許の主分類項目で

あると考えられる。サブコア技術はG05B（制御；調整；制御系または調整系一般；等）、G21C（核物理；核工学、原子炉）、H01L（基本的電気素子；半導体装置；等）といった分野であり、これはコア技術を実現するために必要な処理装置や、または主要適用分野を表している。さらに、周辺技術は、上記以外の応用分野を示していると考えられ、A61B（医学または獣医学；衛生学；診断；手術；個人識別）、G10L（楽器；音響・音声の分析または合成；音声認識）などといった関連分野の広がりを示している（表6）。

表8 対象3分野の代表的要素技術のコア技術と周辺技術

	人工知能分野	液晶分野	バイオテクノロジー分野	
	エキスパートシステム	ドットマトリクス	エンドルフィン	ヒトインターフェロン
コア技術	データ処理技術と(その処理を実現するための)電氣的固体装置	プリンティング技術, ディスプレイなどの表示装置の技術	遺伝子工学	遺伝子工学と有機化学
周辺技術	個人識別, 認識認識などに関する応用技術	広告, サインボード→ゲーム機	装置→飼育・繁殖・飼料	分析・制御方法→医療用具

バイオテクノロジー分野のエンドルフィンは、コア技術が複数存在し、これらの共通する事項を集約すると遺伝子工学であることがわかった。サブコア技術は、開発目的そのものの適用分野である医学または獣医学、衛生学、医療用、歯科用または化粧品製剤などの隣接の技術分野であり、周辺技術は、さらなる応用分野である飼育・繁殖・飼料に関する技術が結果として得られた(表7)。

このほか、液晶技術分野のドットマトリクスに関する技術では、コア技術がプリンティング技術やディスプレイなどの表示装置の技術であり、周辺技術は、さらにその応用である広告、サインボードからゲーム機に関する範囲の技術であるし、ヒトインターフェロンは、コア技術が遺伝子工学と有機化学であり、周辺技術が分析・制御方法から医療用具に至る範囲であるという分析結果が得られた(表8)。

6. 結言

本研究では、データマイニング手法を用いて、

論文と特許文献の概念的な類似性からつながりを示す方法を提示し、人工知能分野、液晶分野、バイオテクノロジー分野の論文について、全特許文献との技術連関分析を行うとともに、本手法の有効性を示した。

データマイニング手法による俯瞰的な分析力を生かして、論文と類似特許の件数について年別推移傾向を表し、論文の隆盛(減少)傾向と特許の傾向が、非常に似ている技術分野が存在することを明らかにした。また類似特許のIPC分類に着目して、各技術に対するコア技術と周辺技術を同定した。現在、この手法を応用したクラスター分析や技術と製品のつながりの分析を実施中であり、基礎技術から製品に至るまでの技術連関の構築を目指している。なお、分析に際して用いたJICSTファイルは、科学技術全般にわたる広い範囲の文献を収容しており、データマイニング分析を用いた技術連関分析を行うにはとても有用であったことを付記する。

参考文献

- 1) 菰田文男. 科学・技術と価値. 東京, 多賀出版, 2000, 288p.
- 2) 菰田文男, 西山賢一, 林 倬史. 情報通信と技術連関分析. 東京, 中央経済社, 1996, 204p.
- 3) 岸田和明. 情報探索の理論と技術. 東京, 勁草書房, 1998, 330p.
- 4) 宮田由紀夫. 共同研究開発と産業政策. 東京, 勁草書房, 1997, 150p.