

特許拒絶理由の統計的分析

—第 36 条を中心に—

Statistical Analysis of Reasons for Refusal of Patents

豊橋技術科学大学情報・知能工学系研究員 **太田 貴久**

PROFILE

2006 年豊橋技術科学大学大学院博士前期課程知識情報工学専攻修了。現在、同大学研究員。修士（工学）、言語処理学会会員、日本知財学会会員。

✉ kikyuu@la.cs.tut.ac.jp

1 はじめに

近年、産業の活性化と国際競争力の強化のために、客観的に分かりやすく、かつ、コンピュータ処理にも適するような日本語（産業日本語）に関する研究・開発・普及活動が行われている。特に、特許文書については、特許版・産業日本語委員会において、様々な取り組みがなされている。その中で、産業日本語による特許文書作成を支援するために、特許請求の範囲（請求項）と明細書の記載方法を定めた特許法第 36 条を類型化し整理する活動が行われている^[1]。

一方、我々は、特許全体の審査プロセスの実態を明らかにするための研究を行ってきた^[2]。本研究では、拒絶理由通知書と呼ばれる文書を大量に収集し、自然言語処理技術を用いて分析を行った。拒絶理由通知書とは、特許審査の結果、拒絶された場合、その詳細な拒絶理由を出願人に知らせる文書である。

特許法第 36 条を類型化し整理するためには、その運用実態を把握する必要がある。そこで、本稿では、これまでの研究 [2] で得られたデータと知見を活用し、特許法第 36 条が実際にどのように運用されているかを把握するための基礎的な分析を行った。

2 分析対象

2.1 拒絶理由通知書について

はじめに、本研究の分析対象である拒絶理由通知書(以

下、単に通知書と呼ぶ)の構造と、分析に利用する情報についての説明を行う。一般的な通知書の構造を図 1 に示す。

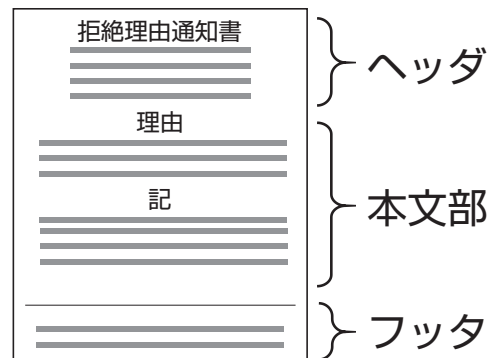


図 1: 拒絶理由通知書の構造

多くの通知書は、図 1 に示すようなヘッダ、本文部、フッタの 3 つの部位で構成される。このなかで、ヘッダは、対象特許の出願番号、適用条文（拒絶の根拠となった条文）など様々な情報が箇条書きで記載されている。一方、本文部は、具体的な拒絶理由が記載された部分であるが、コンピュータで容易に処理できるような定められた書式が存在しない。そのため、本文部から情報を得ようとする場合、何らかの工夫が必要である。

本研究では、図 1 のような通知書について、1) 通知書の対象となった特許の属する技術分野、2) 第 36 条に関する詳細な適用条文、3) 本文部に現れる表現の 3 種類の情報を抽出し、それを分析する。ここで、1) の技術分野は、通知書の対象となった特許の公報に記載されている IPC (International Patent Classification) から、WIPO (World Intellectual Property Organization) による、IPC and PATENT

Concordance Table [3] に基づいて決定した。

また、2) は、ヘッダの適用条文に第 36 条が存在する通知書を対象に、本文部に現れる「第 36 条第○項第○号」というパターンの文字列を抽出し、抽出した文字列の「第○項第○号」をその通知書の詳細な適用条文（以下、詳細条文と呼ぶ）とした。本稿では、特に重要な詳細条文である、第 36 条第 4 項第 1 号（明細書の記載要件）と第 36 条第 6 項第 1～4 号（請求項の記載要件）のみを対象とする。

最後に、3) の本文部に現れる表現とは、本文部に存在する句点で終わる文に現れる表現である。なお、本稿における「表現」とは、文献 [2] と同様に、文中で 1 回以上連続して現れる名詞、動詞、形容詞、助詞「の」と、助動詞「だ」、「ない」とする。

本稿では、第 36 条の運用実態把握のための第一歩として、以上の情報から、通知書を技術分野ごとに分類した際に、1) 各条文が適用される割合に違いはあるか、2) 通知書に用いられる表現に違いはあるか、の 2 点について分析する。

2.2 分析対象について

次に、分析対象とした通知書について説明する。本研究では、2000 年から 2007 年に出願された特許に対する最初の拒絶理由通知書をランダムにサンプリングした 12,773 件の通知書を対象に分析を行う。使用した通知書の一覧を表 1 に示す。これ以降、本研究で用いる図表では、技術分野を表 1 の記号で表す。

3 条文が適用される割合の分析

最初の分析では、技術分野ごとに、第 36 条に関する詳細条文が適用される割合（以下、適用率と呼ぶ）が、どの程度異なるかを確認する。紙面の都合上、すべての技術分野の詳細条文の適用率を載せることは困難である。そこで、本稿では、技術分野ごとに条文の適用率を値とするベクトルを作成し、それらのベクトルに対して主成分分析を行うことで、各技術分野における詳細条文の適用率の傾向を調べる。結果を図 2 に示す。

表 1 使用した拒絶理由通知書

分類	記号	技術分野	通知書数	分類	記号	技術分野	通知書数	
電気工学	E01	電気機械、電気装置、電気エネルギー	1,239	化学	C06	基礎材料化学	527	
	E02	音響・映像技術	1,202		C07	無機材料、冶金	524	
	E03	電気通信	1,028		C08	表面加工	551	
	E04	デジタル通信	312		C09	マイクロ構造、ナノテクノロジー	78	
	E05	基本電子素子	185		C10	化学工学	459	
	E06	コンピューターテクノロジー	2,150		C11	環境技術	357	
	E07	ビジネス方法	1,679		機械工学	M01	ハンドリング機械	599
	E08	半導体	102			M02	機械加工器具	804
機器	I01	光学機器	1,247			M03	エンジン、ポンプ、タービン	546
	I02	計測	814			M04	繊維、製紙	858
	I03	生物材料分析	93			M05	その他の特殊機械	721
	I04	制御	709	M06		熱処理機構	344	
	I05	医療機器	453	M07		機械部品	3,074	
化学	C01	有機化学、農業	918	M08		運輸	872	
	C02	バイオテクノロジー	143	その他	001	家具、ゲーム	544	
	C03	製薬	542		002	その他の消費財	393	
	C04	高分子化学、ポリマー	597		003	土木技術	590	
	C05	食品科学	237					

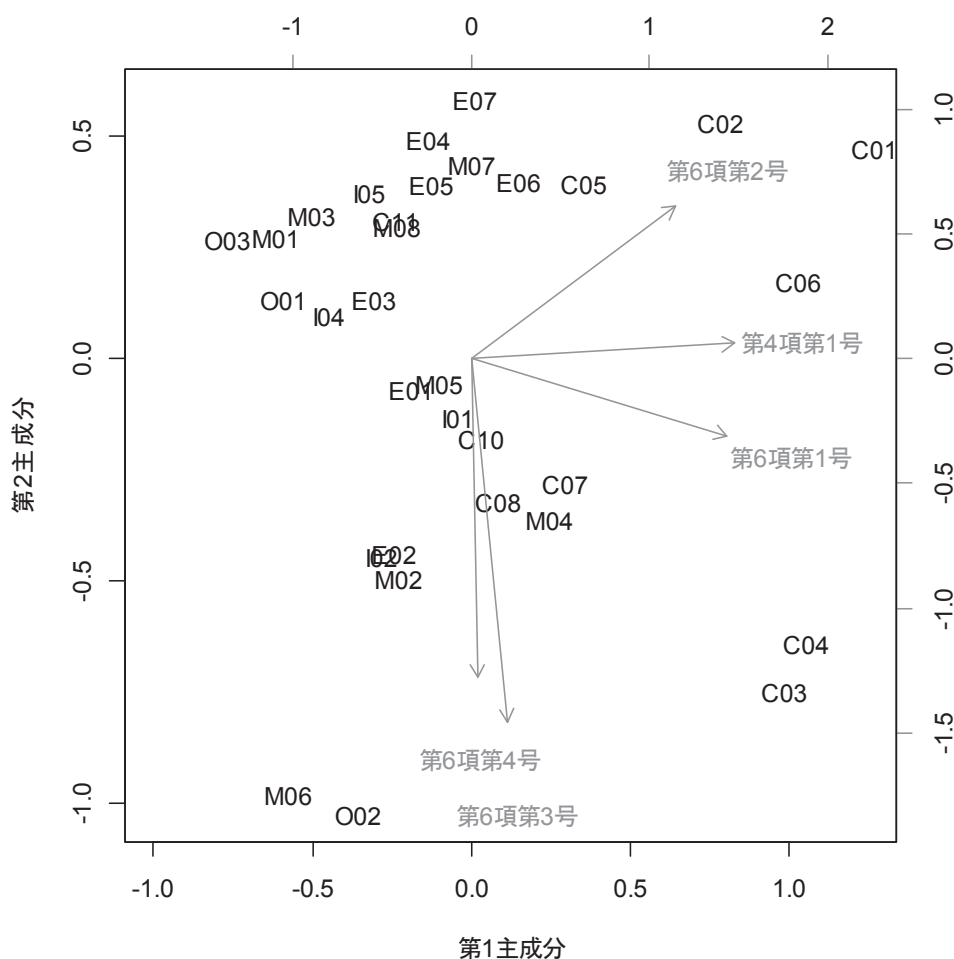


図2 詳細条文適用率の主成分分析（技術分野）

図2は、第1主成分と第2主成分にもとづいて、各技術分野をプロットした図である。なお、第2主成分までの累積寄与率は0.75である。

図2のように、各技術分野は、明確なクラスター（グループ）を作らず、分散して配置されている。そのなかで、化学分野（Cnn）は、図2の比較的右周辺部に配置されている。次の表現の分析では、通知書を化学分野とそれ以外に分けて分析する。

4 通知書に用いられる表現の分析

次に、通知書に用いられる表現についての分析を行った。第36条第○項第○号違反という情報のみでは、その具体的な拒絶理由は分からない。例えば、第36条第6項第1号違反だけでは、「請求項に現れる用語がどこ

にも定義されていない」や、「実施形態が1つ以下しか記載されていない」といった詳細な拒絶理由は分からない。そこで、具体的な拒絶理由を調査するために、通知書の本文に用いられる表現の調査を行った。本稿では、前節における、詳細条文の適用率の分析結果より、化学分野とそれ以外の分野における表現の違いを確認する。具体的には、ある条文に関する記述中に、どのような表現が用いられやすいかを、化学分野とそれ以外の分野間で比較する。

4.1 表現の出現確率と特徴量

本調査では、各詳細条文 a に関する記述において、ある表現 w が用いられる確率 $P(w|a)$ を比較する（以下、これを表現の出現確率と呼ぶ）。この確率を単純に求めようとした場合、以下の式のように求めることが考えられる。

条文 a が適用され、かつ、表現 w が用いられた通知書数
 条文 a が適用された通知書数

しかしながら、上式によって求められる確率は、「条文 a が適用される通知書において、表現 w が用いられる確率」である。これに対して、本研究で用いる確率は、「通知書の条文 a に関する記述において、表現 w が用いられる確率」である。これらは似ているが大きく異なる。前者の確率には、共に適用される条文の影響がそのまま含まれてしまう。そこで、本研究では、ある条文 a の記述中に表現 w が用いられない確率をもとに、表現 w が用いられる確率を推定する。

ある表現が、条文 a が適用される通知書で用いられない確率を \bar{p}_a としたときに、収集した通知書が観測される確率（尤度）は以下の式のように求めることができる。

$$L(\bar{p}_1, \bar{p}_2, \dots) = \prod_A \binom{n_A}{n_A - f_A} \bar{p}_A^{(n_A - f_A)} (1 - \bar{p}_A)^{f_A}$$

$$\bar{p}_A = \prod_{a \in A} \bar{p}_a$$

ここで、 A は条文の適用パターンを表し、 n_A は A というパターンで条文が適用された通知書の数を、 f_A は、その中で対象とする表現が現れた通知書の数を表す。

上式を最大化するような、 \bar{p}_a を求めることで表現の出現確率を求める。なお、この確率の求め方に関しては、文献 [2] にその詳細を記載している。

以上のように求めた表現の出現確率が高いものが、その拒絶理由の傾向をよく表す表現かと言うとそうではない。例えば、動詞「する」は、様々な通知書に高頻度で現れるため、表現の出現確率は高くなる。しかしながら、これは拒絶理由の傾向を表すような表現ではない。本研究では、このような一般的な表現を除くために、表現の特徴量を用いる。ここで、条文 a に対する表現 w の特徴量 $S(w|a)$ を以下のように定義する。

$$S(w|a) = P(w|a) \log \left(\frac{|A|}{\sum_{i \in A} P(w|a_i)} \right)$$

ここで、 A は分析対象の条文の集合を表す。本特徴量は、情報検索や文章要約といった自然言語処理の分野で用いられる、古典的な語の特徴量（重み）を表す指標である tf-idf (term frequency - inverse document

frequency) に基づいている。tf-idf では、ある文書における語の特徴量について、その文書で数多く現れ、かつ、他の多くの文書で現れない語ほど大きな値を与える。すなわち、助詞「は」や動詞「する」の tf-idf は低くなる。tf-idf については、文献 [4] 等の自然言語処理のテキストに更に詳しい解説がある。

4.2 表現の比較

第 36 条の中でも特に拒絶の理由とされやすい、第 36 条第 6 項第 1 号（サポート要件）、第 36 条第 6 項第 2 項（明確性要件）と、第 36 条第 4 項第 1 号（実施可能要件）について、化学分野とその他の分野のそれぞれ、特徴量が高い表現 20 個を表 2、表 3 に示す。

表 2 と表 3 を比較すると、特徴量の高い表現は 2 つの分野間で大きく異なることが確認できる。例えば、第 6 項第 1 号において、「一般化」という表現が、化学分野では、その他の分野に比べ、高い確率で現れる。これは、化学分野では、「出願時の常識に照らしても、請求項に係る発明の範囲まで、発明の詳細な説明に開示された内容を一般化できるとは言えない」ことで拒絶される場合が多いことが原因であると考えられる。

5 おわりに

本稿では、大量の拒絶理由通知書から、特許法第 36 条の運用実態を把握するための基礎的な分析を行った。その結果、化学分野とその他の分野では、第 36 条の運用に違いがあることを確認した。

今後、技術分野を細分化して分析を進めるとともに、単純な表現の出現確率以外（例えば、文法パターン）も用いた詳細な分析を行う予定である。

表2 特徴的な表現（化学分野）

第36条第6項第1号			第36条第6項第2号			第36条第4項第1号		
表現	確率	特徴量	表現	確率	特徴量	表現	確率	特徴量
拡張	0.288	0.950	不 明確	0.430	1.139	できる 程度	0.717	1.105
一般 化	0.280	0.933	引用 文献 #	0.522	0.752	十分	0.745	1.043
技術 常識	0.327	0.803	事項	0.407	0.735	実施 する	0.548	1.032
開示	0.318	0.786	不明	0.262	0.729	実施 する こと	0.514	1.026
化 できる	0.195	0.760	意味	0.202	0.709	説明 の 記載	0.965	1.017
一般 化 できる	0.191	0.750	引用 する	0.210	0.673	出願 の 発明 の 詳細	0.394	0.961
開示 する	0.281	0.745	新た	0.231	0.667	実施 例	0.352	0.616
いえる	0.352	0.731	ため	0.342	0.659	明細	0.510	0.603
出願 時 の 技術 常識	0.216	0.717	特定 する	0.254	0.658	明細 書	0.467	0.586
具体的	0.464	0.689	「～」	0.890	0.654	実施 できる	0.156	0.583
本願 発明	0.309	0.665	行う	0.319	0.652	本願 明細	0.247	0.565
例	0.489	0.661	# の 「～」	0.185	0.645	構造	0.258	0.556
効果	0.406	0.657	発見 する	0.214	0.643	化合 物	0.362	0.550
実施 例	0.371	0.648	通知 する	0.212	0.641	以外	0.287	0.547
ない	0.332	0.642	よう	0.573	0.636	方法	0.310	0.545
特定	0.386	0.608	拒絶 の 理由	0.209	0.634	化学	0.193	0.537
超える	0.211	0.603	該当 する	0.427	0.631	製造	0.286	0.532
同様	0.339	0.598	下記 の 刊行 物	0.658	0.626	本願 明細 書	0.218	0.522
上記	0.477	0.597	基く	0.652	0.619	- #	0.362	0.517
課題	0.175	0.590	発明 の 属する 技術 の 分野	0.656	0.617	# -	0.375	0.514

表3 特徴的な表現（その他の分野）

第36条第6項第1号			第36条第6項第2号			第36条第4項第1号		
表現	確率	特徴量	表現	確率	特徴量	表現	確率	特徴量
課題	0.343	0.774	不 明確	0.346	0.812	説明 の 記載	0.912	1.285
解決	0.299	0.772	意味	0.264	0.672	出願 の 発明 の 詳細	0.581	1.235
反映 する	0.199	0.730	適宜	0.289	0.669	できる 程度	0.550	1.186
解決 する	0.265	0.729	周知	0.351	0.659	実施 する こと	0.515	1.172
化	0.245	0.680	発見	0.347	0.655	十分	0.543	1.121
解決 する ため	0.178	0.677	通知 する	0.352	0.654	段落	0.399	0.811
ため の 手段	0.165	0.651	発見 する	0.347	0.654	明細	0.360	0.695
する ため の 手段	0.160	0.645	新た	0.352	0.654	明細 書	0.353	0.689
含む	0.288	0.637	拒絶 の 理由	0.348	0.651	例	0.275	0.681
拡張	0.152	0.628	装置	0.327	0.650	段落 (~)	0.242	0.669
発明 の 課題	0.163	0.627	引用 する	0.260	0.643	経済	0.160	0.639
解決 する ため の 手段	0.150	0.623	引用 文献 #	0.629	0.636	経済 産業 省令	0.157	0.631
開示 する	0.203	0.619	構成	0.473	0.634	定める ところ	0.157	0.631
一般 化	0.146	0.610	不明	0.415	0.625	図	0.258	0.617
対応	0.264	0.594	事項	0.471	0.605	技術 的	0.292	0.608
いえる	0.191	0.568	該当 する	0.267	0.604	等	0.375	0.603
認める	0.355	0.557	行う	0.398	0.603	不明	0.392	0.590
化 できる	0.121	0.554	特定	0.310	0.602	意義	0.176	0.583
一般 化 できる	0.120	0.549	出願 前 日本 国内	0.209	0.602	図 #	0.210	0.578
対応 する	0.221	0.541	# の 「~」	0.177	0.588	本願	0.581	0.578

参考文献

- [1] 松田成正, 谷川英和. 特許版・産業日本語の活動報告. 第5回産業日本語研究会・シンポジウム予稿集, pp.15-39. 2014.
- [2] 太田貴久, 野中尋史, 平岡透, 増山繁. 自然言語処理を用いた特許拒絶理由の統計的分析と審査プロセス解析への応用. 日本知財学会誌, Vol. 10, No.3, pp.61-78. 2014.
- [3] WIPO. IPC and Technology Concordance Table. http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=117672. 2009.
- [4] 北研二, 津田和彦, 獅々堀正幹. 情報検索アルゴリズム. 共立出版. 2002.