

専門的なデータ分析ツールに基づく 中国專利データの高度な分析方法

Chinese Patents Analysis based on Deep Data Mining



知識産権出版社有限責任公司 研究發展部プロダクトマネージャー

張 秋懿

ソフトウェアエンジニア/中国国家知識産権局中堅人材。主に INTEcovery 專利ナビゲーション分析システムと專利ソナーおよびその派生製品の研究開発を担当する



知識産権出版社有限責任公司 国際部副主任

傅 毅冬

法学修士/研究員補。主に特許データの加工・マイニングプロジェクト及び諮問・調査プロジェクトに関する国際協力及びリソースの収集を担当する

1 はじめに

專利データの統計分析は、知的財産サービス業界で重要な業務内容の1つである。従来の專利データ統計分析は、検索して得たデータに対して、データのクリーニング、索引作成を行う作業である。アナリストは多くの時間と労力を費やさねばならず、従来のデータ分析機能では專利分析が同質化しがちであり、またデータモデルが限られているため、アナリストの能力の発揮に足かせがかけられる。ビッグデータ、人工知能技術の幅広い応用に伴い、專利データ分析ツールは大きな進歩を遂げ、專利アナリストに多くの手助けをしている。本稿では、分析例を用いて、次の3つの側面からこれについて検討したい。

- 1) スマートな索引技術が、專利アナリストの作業時間を短縮させる
- 2) 特色のある分析機能で、高度な專利分析を実現する
- 3) 自由度の高い分析モデルの組み合わせにより、專利アナリストによる分析において個人的特徴が生まれる

2 スマートな索引技術が、專利アナリストの作業時間を短縮させる

従来の索引作成では、專利アナリストが專利を1件ずつ読み、專利データの内容に基づいて、人力で索引を作成しなければならない。一方、ビッグデータルール又は人工知能技術に基づいてスマートな索引を行うことで、アナリストの作業負荷を軽減することは可能である。

1. 自動化データルールによる「眠っている」專利の発見

「眠っている」專利の基本要件は次のとおりである

- 1) 專利権者が中国の大学又は研究機関であること
- 2) 專利権付与後に3年間以上存続していること
- 3) 專利の収益化に関する挙動がないこと

言うまでもないが、下図に示すように、このような特定のデータ索引ルールは分析ツールから自動的に提供できる。

自動化ルールの強みは、研究開発の主体、存続期間、專利の状態、価値範囲についてルールを追加又は変更することで、專利アナリストにより多くの自由度と可能性をもたらすことができるということにもある。

図 1

当方は、当該ルールを用いて、「パワーバッテリーパック管理モジュール」技術関連の 1 万 1974 件の中国専利に対して自動的索引を行った。当該ルールに適合する「眠っている専利」が 319 件得られ、保有機関の分布は次のとおりである。

表 1

機関	眠っている専利の数
清华大学	14
江蘇大学	11
北京理工大学	11
華南理工大学	10
浙江大学	8
上海理工大学	7
吉林大学	7
上海空間電源研究所	6
上海交通大学	6
西安交通大学	6

以上から分かるように、このような専利の保有者をすぐに特定し、後には収益化に向けて関連の作業を行うことができる。

2. 自動化ルールに基づくコア（障壁）専利のマイニング

これが上記と同様の課題である。コア（障壁）専利とは、審査官 X 又は Y が、一定の数の専利に対し授権をしないために引用文献として使う専利のことである。このような専利の存在は、他の専利の授権に直接的な影響

を与えるため、技術的にはコア的なもの、又はその保有者が技術障壁を築くために役立つものと見なすことができる。

また、ツールの中では、下図のようにルールを設定することができる。

図 2

授権を阻む専利の数はいったん 3 件と設定し、依然として上記の「パワーバッテリーパック管理モジュール」技術関連の専利を例として、自動的索引を行った。

表 2

機関	コア（障壁）専利数	専利総数ランキング
奇瑞汽車股フン有限公司	6	16 位
比垂迪股フン有限公司	5	1 位
華南理工大学	4	14 位
中興通訊股フン有限公司	4	69 位
惠州市億能電子有限公司	3	13 位

以上から分かるように、分析の結果が、業界の技術競合状況の分析及び判断に対し、新しい見方を提供している。

3. 複雑なビッグデータモデルに基づく専利価値の等級付け索引

弊社・知識産権出版社は、長年にわたる技術の蓄積を経て、32 項目のデータ指標に基づく専利価値評価ビッグデータモデルを構築しており、それは 100 点制で専利に点数を与え、その価値を反映する仕組みになっている。当方は、データ分析に際し、当該ビッグデータモデルによる点数付与の結果を踏まえて、データの等級付け索引を行うことができる。図 3 を参照されたい。

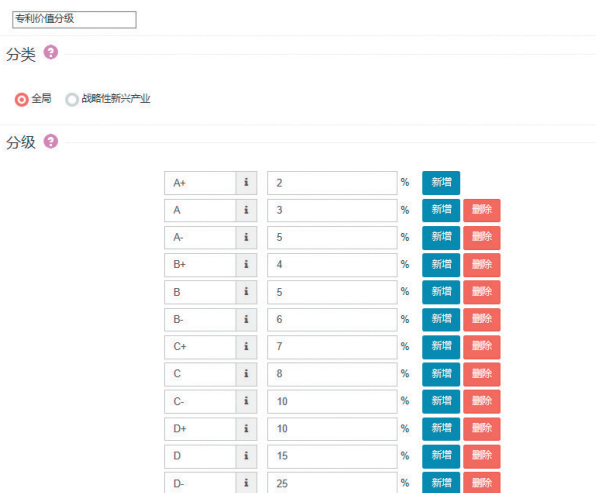


図 3

当方は、「パワーバッテリーパック管理モジュール」技術関連の専利を A+ から D- まで 12 級に分けていた。異なる級の専利データの分布を「ピラミッド型」の分布として設計した。即ち、全ての A 級専利が総数の 10% を、B 級専利が総数の 15% を、C 級専利が総数の 25% を、D 級専利が総数の 50% を占めている。

このように設計し自動的索引を行った後に、業界トップ 5 位の出願人に対し、その実際の専利価値の分布状況が専利の出願量に合致するかどうかを検討することができる。

当然ながら、異なるデータ分布のタイプを設計して、分析の方式を調整することができる。

4. セマンティックベースのスマートな技術で、効果行列による分析を実現する。

分析ツールは、スマートな方式により、イノベーションを行う専利の主体及び改善効果を直接抽出することができる。効果行列による分析を自動的に生成して、技術の研究開発の方向性に関する研究に手助けをする。

電池組	5	8	9	4	1	2	1	4	4	2	3	4	1	2	1	3	2			
電池	3	4	5	4	2	1	2	3	1	5	2	2	3	2	3	3				
太陽電池	2	2	4	1	1	1	1	1	1	2	1	1	1	1	1	1				
小型電池	2	7	10	6	3	7	8	2	2	7	3	2	2	7	2	2	5	1		
電池回路	6	2	1	5	1	3	2	3	1	1	2	1	2	2	2	2	2			
充電	8	7	1	5	8	6	2	5	2	6	1	2	1	1	4	1	3			
電池組	4	9	1	8	5	2	5	7	3	1	2	1	4	4	2	3	2	1	3	
熱管理	15	8	26	2	1	11	4	2	2	12	2	11	2	2	9	4	10	1	4	
自動車	11	16	11	16	11	4	7	2	7	5	2	2	2	2	5	3	3	6		
電池	35	23	22	17	22	25	13	11	13	13	18	10	6	12	8	10	11	10	6	8
	耐久性 提升	安全 提升	温度 提升	成本 提升	電圧 提升	効率 提升	穩定性 提升	成本 下降	工序 提升	散熱 提升	電容量 提升	利用率 提升	電流 提升	防護性 提升	熱効率 提升	構造 提升	電圧 提升	熱能 提升	信頼性 提升	可溶性 提升

図 4

3 特色のある分析機能で、高度な専利分析を実現する

1. 技術競合分析

当方は依然として「パワーバッテリーパック管理モジュール」技術関連の専利を選択し、ツールに搭載されている専門的な分析機能「技術競合分析」を利用して、「A社」および「B社」両社を選んで技術競合の比較を行ったところ、次のことがわかった。

表 3

	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-
C社	6	12	14	8	14	15	16	15	14	7	8	5
D社	1	3	12	11	6	10	19	10	18	18	8	7
E社	0	5	4	5	5	5	7	12	12	3	6	7
F社	4	5	8	2	4	4	12	9	8	2	0	1
G社	1	0	1	0	1	3	6	10	13	12	8	2

また、ツールは自動的ビッグデータ分析を提供する。

「【A社】が競合に参入している専利は少なくとも【118】件あり、【A社】の専利データ全体の約【88%】を占めている

【B社】が競合に参入している専利は少なくとも【58】件あり、【B社】の専利データ全体の約【98%】を占めている

【A社】は【3】つの技術項目において専利数が優位に立ち、【B社】は【1】つの技術項目において専利数が優位に立ち、両者は【1】つの技術項目において件数が同じである

これに対して、【A社】は少なくとも【8】つの独自の技術項目を持ち、関連専利が【9】件で、実体【A社】の専利データ全体の約【7%】を占めている

【B社】は少なくとも【0】の独自の技術項目を持ち、

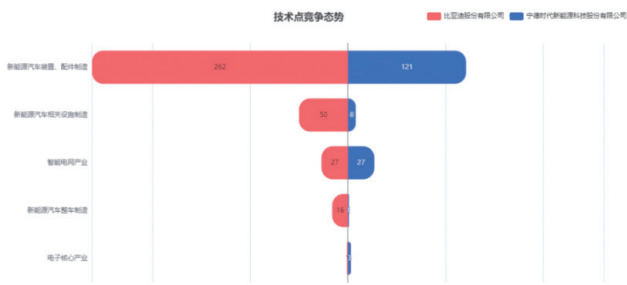


図 5



図 6



図 7

関連特許が【0】件で、実体【B社】の特許データ全体の約【0%】を占めている

こうした機能により、技術競合における両社の長所と短所およびそれぞれの重点分野が一目瞭然になった。

2. 新規参入企業分析

当方は依然として「パワーバッテリーパック管理モジュール」技術関連の特許を選択し、ツールに搭載されている専門的な分析機能「新規参入者分析」を利用している。「2000年～2005年」「2006年～2010年」「2011年～2015年」「2016年～2023年」の4つの期間を設定した。次の結果を得た。

当該設定において、出願人の総数は【7365】個である

2000～2005年では、出願人の総数は【162】個であり、新規参入者は【162】個である

2006～2010年では、出願人の総数は【611】個であり、新規参入者は【589】個である

2011～2015年では、出願人の総数は【1824】個であり、新規参入者は【1687】個である

2016～2023年では、出願人の総数は【5305】個であり、新規参入者は【4892】個である

上記の分析により、「パワーバッテリーパック管理モジュール」技術関連の技術競合の特徴について高度な分析を行うことができる。少なくとも次の結論を得ていた。

- 1) 当該分野では技術革新を行う者が爆発的に増えている
- 2) 新規参入企業は成長が極めて早く、5年が経たないうちに特許技術の数が優位に立っている
- 3) 中国国内企業の特許の割合が高くなりつつある

4 自由度の高い分析モデルの組み合わせにより、特許アナリストによる分析において個人的特徴が生まれる

分析ツールは、カスタム分析においてより多くの統計項目フィールドを利用することができる。

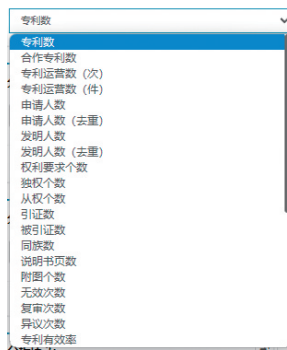


図8 より多くの分析フィールド



図9

また、全ての索引フィールドを用いるカスタム分析の組み合わせが可能になる。



図10

このような組み合わせでより多くの分析モデルを得ることができ、特許アナリストはその中から分析モデルを選択して個人的特徴のある分析を作成できる。

例えば、当方は「A社」の「パワーバッテリーパック管理モジュール」技術関連の特許に対し、異なる分析の組み合わせにより独特なデータプロファイルを設計することを試みた。

当方は出願人から分析するという従来のやり方を捨て、「A社」が「現在の第1権利者」である特許に対す

る分析を試みた。次のデータプロファイルを得ていた。

表4

A社	
データ項目	値
特許総数	128 (件)
特許数	84 (件)
特許授權率	69.05%
有効特許数	90 (件)
協力特許の割合	0%
平均請求項数	14.24 (項)
平均明細書頁数	9.97 (頁)
発明者総数	255 (人)
平均存続期間	4.67 (年)
平均被引用回数	4.58 (回)
コア(障壁)特許数	3 (件)
拒絶査定不服審判回数	1 (回)

以上から分かるように、我々は、単純な分析フィールドの組み合わせにより、企業が現在特許権を持っている特許に対して高度なデータプロファイリングを行い、そして特許総数、特許の質、特許収益化、チームメンバーなど複数の側面から対象企業に対し様々な視点から踏み込んだ分析を行うことができる。

5 おわりに

特許データの高度なマイニング分析は、果てしなく続いていく作業である。時代の変化に伴い、特許データの内容が増え、データ抽出作業の手段が進化していき、異なるデータフィールド間の関連性にも変化が生まれるだろう。高度な特許分析ツールを利用すれば、特許アナリストはそのノウハウを一層発揮することができ、そして分析を行って膨大なデータの中から技術競合又は市場の動向に関する高度な情報を見つけられると考えられる。

