

# 中国特許出願動向分析

## 中国語テキストマイニングを用いた原文分析

○袁方<sup>1)</sup>, 小林義典<sup>1)</sup>, 安藤 俊幸<sup>2)</sup>

松下電器研究開発（蘇州）有限公司<sup>1)</sup>, 花王株式会社<sup>2)</sup>

〒215123 中国江蘇省蘇州市工業園区鐘南街300号

Tel: +86-512-62584710

E-mail: kobayashi.yoshinori@jp.panasonic.com

## Analysis of IP application trends in China

### Chinese patent analysis using text-mining method

YUAN Fang<sup>1)</sup>, KOBAYASHI Yoshinori<sup>1)</sup>, ANDO Toshiyuki<sup>2)</sup>

Panasonic R&D Center Suzhou Co.,Ltd. <sup>1)</sup>, Kao Co.,Ltd<sup>2)</sup>

NO.300 Zhongnan Street,Suzhou Industrial Park,Suzhou 215123,P.R.China

Phone: +86-512-62584710

E-mail: kobayashi.yoshinori@jp.panasonic.com

#### 【発表概要】

近年世界最大の出願件数を有する中国特許文献を網羅的に調査することは、現地企業を中心とする中国市場での技術動向を把握する上で重要な意味を持つ。しかしながら、膨大な件数の特許情報の確認、分析に対して時間と工数の確保が課題となることは言うまでもない。

特許情報の内容を機械的に把握する手段としてテキストマイニングが有効であることは周知である。しかし、中国特許に対して有効に分析でき、かつ普及しているツールはまだ少ないのが現状である。

本報では中国特許に対して、中文環境下でテキストマイニング手法を用いて精度よく分析できる環境を模索する。特に、その中でもテキストマイニングのタームの抽出段階で極めて重要な「分詞（分かち書き）」に着目し、いくつかの入手可能なツールを比較評価するとともに、現状の課題も明らかにする。

加えて、選定したツールを用いて空調分野での特許を分析し、出願動向などの把握を試みる。

#### 【キーワード】

中国特許, テキストマイニング, Macro 分析, 統計分析, 分詞, 分かち書き

## 1. はじめに

多量の特許情報の分析に関し、日本では既に和文、英文での「テキストマイニング（以下、表中等で「TM」と記載する）」を用いた分析技術が多用されており、またその用に供するツールも有償、無償含めて多数存在している。

中国特許に対して同様の分析をするには DWPI (Thomson Reuters) や CNIPR (知識産権出版社) 等で提供されている英文抄録などを利用して、既存のツールを用いて分析する手法もあるが、全文分析する場合等、翻訳精度の課題の存在も否めず、中国語原文による分析が必要である。

しかしながら、中国語の特許データを分析するためのテキストマイニングツールは現時点筆者の知る限りではまだ十分に開発されておらず、その活用も進んでいないのが現状である。

本報では、中国語ベースでのテキストマイニング及び解析を行うことができるツールの模索とその評価を行い、利用可能なツールを選定し、事例として空調分野での分析を行った。

## 2. 中国語分析ツールの選定

### 2.1 テキストマイニングツールの模索

日本語は優秀な形態素解析器が多数あり、また英語は単語が分かれているのでテキストマイニングに適応しやすいが、中国語は形態素分詞が難しい(図1)。

なお、中国のテキストマイニングでは形態素の「分かち書き」は一般に「分詞」と呼ばれており、本稿ではあえて「分詞」と記載している。

既存の分詞ツールを評価する上で、筆者は分析する際に必要な「出現頻度の統計」と「フィルターの設定機能の

有無」が重要であると考えている。

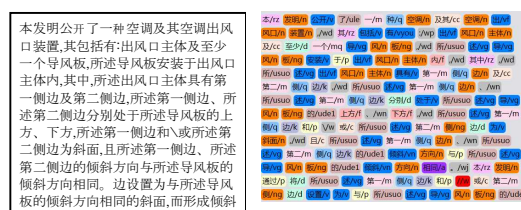


図 1. 中国語の分詞

先ず中国語テキストマイニングが可能なツールを抽出し、検証を行った。現状、中国語形態素分析が可能なツールはまだそれほど多くないが、研究施設と大学から開発されたフリーソフトがいくつかあり、WEBサイト、文献などを通じて抽出した下記の中国語分詞ツールについて評価を行なった。対象は MTMineR(MLTP)<sup>1)</sup>、NLPIR(ICTCLAS)<sup>2)</sup>、Rostnat<sup>3)</sup>、PatentMap EXZ の中国語版 Tecres<sup>4)</sup>、IKAnalyzerNet<sup>5)</sup>、IKAnalyzer のエンジンを利用した PatAnalyzer<sup>6)</sup>の、以上 6 件である(表 1)。

ソフト名	出所	TM	頻度統計	フィルター
MTMineR (MLTP)	同志社大学 金明哲教授	○	△ 難	△ 難
NLPIR (ICTCLAS)	中国科学院	○ 精度高	×	×
RostCM (Rostnat)	武汉大学インターネット 科学研究センター	○	○	○
Tecres (PatentMap EXZ)	インパテック社	△ 精度低	○	×
IKAnalyzerNet	林 良益	○	×	×
PatAnalyzer (IKAnalyzerNet)	安藤 俊幸	○	○	○

表 1. TM ソフトの比較結果

この中から、①中国語のテキストマイニングが可能、②頻度統計とフィルター機能を有するツール 2 件を選定し、出願番号 201110176853.7 の要約を用いて分詞の精度を検証した。



図 2. Rostnat と PatAnalyzer の TM 事例

その結果、Rostnat と、PatAnalyzer が分詞精度と、テキストマイニングでの統計分析においていずれも活用性が高いことがわかった (図 2)。

一方、Tecres は統計分析に優れたツールだが、分詞時の単語の切り分けが長目であるため統計に向かず、またノイズ除去もユーザ側でカスタマイズできないことから、今回の評価では分析利用に適さないと判断した (図 3)。

No.	データ名称	所有
1	温度差	1
2	外側	1
3	凝露現象	1
4	空調	1
5	空調出风口装置	1
6	迎風区	1
7	室内温度分布	1
8	斜面	1
9	出风口主体	1
10	出风口装置	1

図 3. Tecres の分詞結果

## 2.2 統計分析及び可視化ツール評価

中国語の可視化での最大の問題は「文字化け (乱码)」であり、中国語が正常に表示できるツールの中から、欧米の研究機関によって開発されている Cytoscape<sup>7)</sup>、Gephi<sup>8)</sup>、Ucinet6<sup>9)</sup> と、Rostnat の四つのツールに対して評価を行った (表 2)。

ソフト名	出所	データ分析	可視化
Cytoscape	U.S. National Institute of General Medical Sciences	×	○
Gephi	Gephi team	×	○
RostCM (Rostnat)	武漢大学インターネット科学研究センター	○	○
Ucinet6 (Netdraw)	Analytic Technologies	○	○

表 2. 可視化ソフトの比較結果

結果、Rostnat と Ucinet は分析から可視化の作業が一貫して完了できることから最も適切であると評価した。

また、クラスター分析ができる Ucinet を除いて、現時点で中国語原文の可視化は、「共起ネットワーク」程度しか表示できないため、多次元尺度法等の可視化ツールの探索が今後の課題と考える (注: プログラミング言語「R」などを用いて作成する場合を除く)。

Cytoscape と Gephi はデータ分析は不可だが、事前に手動でデータ処理を行えば可視化できることも判明した。

## 2.3 解析ツールの選定

以上の結果から、①「Rostnat」、または②「PatAnalyzer+Ucinet6 の組み合わせ」の二つの方式で中国語原文のテキストマイニング、統計分析、可視化の作業が比較的精度よく、かつ簡単に実施できるということを確認した。

## 3. 分析事例

昨今、中国の空調分野では主要メーカーを中心に出願件数の増加が顕著であり、出願状況の概要把握が期待される技術分野のひとつである。

そこで、中国の近年の空調分野での出願トレンドを把握すべく、テキストマイニングで特許出願の全体像及び主要メーカーの動向分析を試みた。

### 3.1 特許の抽出

検索条件としては JPO の平成 26 年度の中国エアコン分野の特許出願技術動向調査報告書<sup>10)</sup>に記載の調査範囲を参考に、2015 年 6 月 30 日までの、空調分野に関連する特許 32,333 件を抽出した (表 3)。また、分析対象のテキストは「要約 (摘要)」に対し、PatentSquare<sup>11)</sup>を用いて抽出した。

式No.	登録件数	検索項目	条件式
S001	34,134	IPC	F24F 01/00+F24F 11/00+F24F 13/08+F24F 13/10+F24F 13/14+F24F 13/15+ F24F 13/16+F24F 13/18+F24F 13/20+F24F 13/22+F24F 13/24+F24F 13/26+ F24F 13/28+F24F 13/30+F24F 13/32
S002	6,110,428	公告・登録	:20150630
S003	4,730,809	公開日	:20150630
S004	9,155,402	論理式	S002+S003
S005	32,333	論理式	S001*S004

表 3. 空調分野抽出用検索式

### 3.2 空調分野出願全体像の分析

図 4 に出願件数の推移を示した。特に 2008 年以降顕著に増加し、現在までの 7 年間に 24,815 件もの多数の出願が出されており、2014 年は 6,059 件に達している。このような多量の出願に対しテキストマイニングのような手法で概要を把握することは非常に有効である。

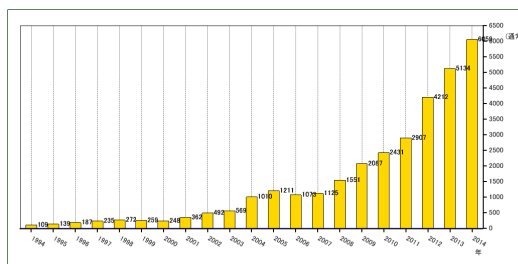


図 4. 中国空調分野の出願件数推移

今回は選定したツールのうち Rostnat を用いて 32,333 件の特許の要約の分析を実施し、中国空調分野の技術分布を把握した (図 5)。同図において、ノードはテキストマイニングで抽出された単語を示し、かつその大きさはその単語の頻出度と関連する単語との接続の多さを示す。またリンクは単語間の関連性を示し、その太さはその関連の頻出度を示している。

ちなみにこの数のデータ (漢字 7,445,774 文字) に対して、Rostnat でのテキストマイニング分析は標準のデスクワーク仕様の PC (intel i5 3320 2.6GHz/RAM 4GB) を用いて 205 秒程度で完了し、統計処理に 170 秒、フィルタリングに 3 秒、描画には 32 秒程度で処理完了した。

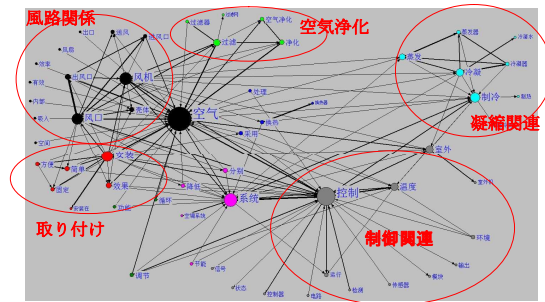


図 5. 全件の TM 結果

また、2014 年の出願に対する分析の結果と比較すると、全体的な分布は変わらないものの、「空気浄化分野の拡大」と「スマート機能関連の特徴語」が多くなっていることが分かった (図 6)。

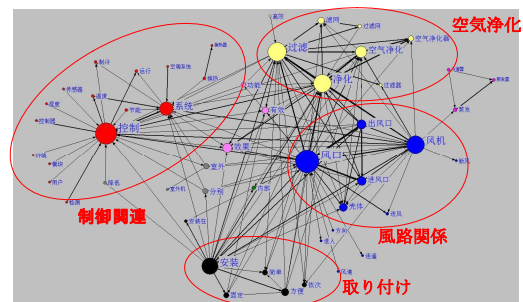


図 6. 2014 年の TM 結果

### 3.3 中国主要メーカーの分析

次に、販売台数総計が中国の空調市場シェアの 6 割を越える主要メーカー三社<sup>12)</sup> (A 社、B 社、C 社) の出願動向の統計分析を行なった (図 7)。加えて三社ごとの出願内容についてもそれぞれ分析を実施した (図 8~10)。

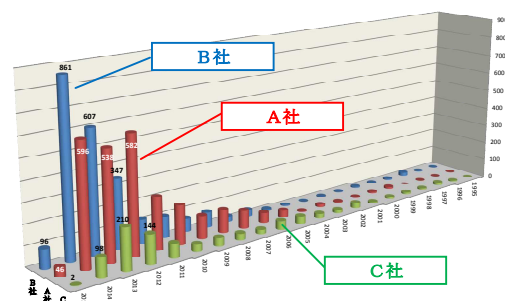


図 7. A/B/C 社の出願件数の年間推移

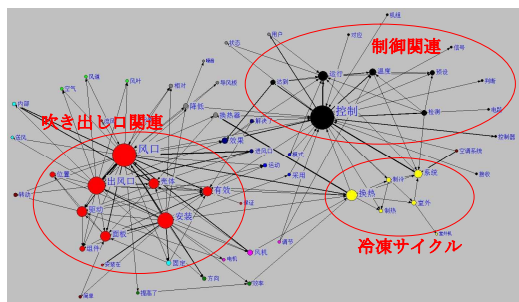


図 8. A 社の TM 結果

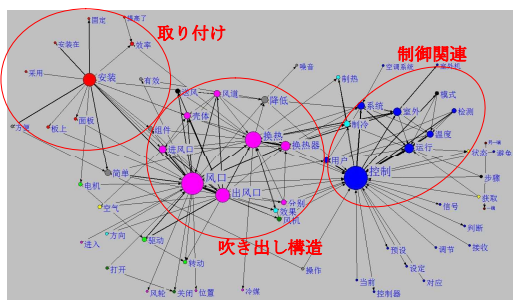


図 9. B 社の TM 結果

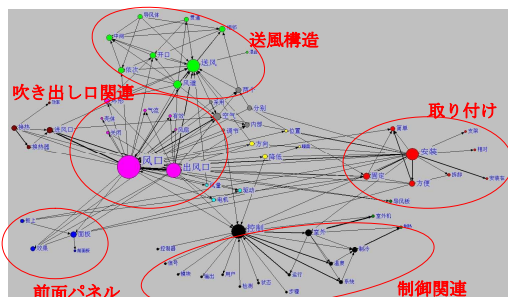


図 10. C 社の TM 結果

その結果、三社ともに「吹き出し」関連の特徴語が多く見られ、中国での出願は「吹き出し口の配置・構造や風向羽根の動き」に注力されていると推定できる。

三社の出願に対し、要約にキーワード (KW) 「風口」が含まれる案件を抽出してみたところ、「風口」にかかわる出願が3社とも3割ぐらゐを占めることがわかった (表4)。

	抽出件数	全件数	割合
A 社	750	2666	28%
B 社	725	2400	30%
C 社	237	869	27%

表 4. KW 「風口」での抽出結果

また、32,333 件のうち、「風口」に関連する特許を検索すると、7,802 件が

抽出され、これをテキストマイニング分析した結果、KW 「風口」は「空気浄化」との関連性が高いことが確認できた(図 11)。業界全体も含め、近年この分野の出願が増えているものと推定できる。

### 3.4 中国企業と日本企業の特徴対比

次に、中国メーカーA社と日本メーカーD社の出願について比較した(図 12~14)。

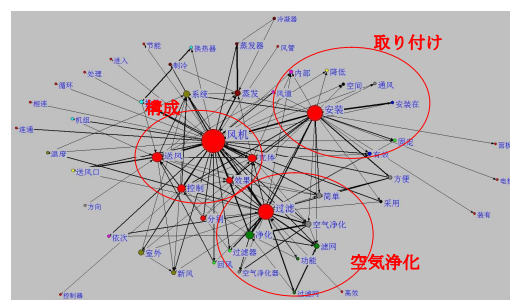


図 11. 「風口」の TM 結果

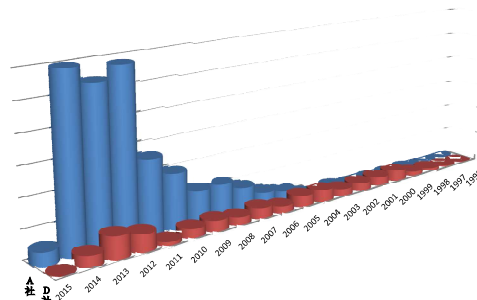


図 12. A 社と D 社の出願推移

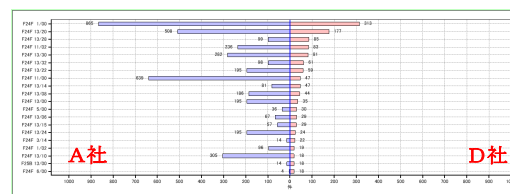


図 13. A 社 vs D 社 IPC 別出願件数

D 社の出願の特徴語を抽出(図 14)し、図 8 の A 社の分析結果と比較したところ、A 社と D 社では「制御」と「冷凍サイクル」は共通する一方で、A 社では「吹き出し口の構造」と「取り付け位置」に注力されており、D 社では「冷媒」、「空気浄化」と「風向制御」に対して注力されていることがわか

った。

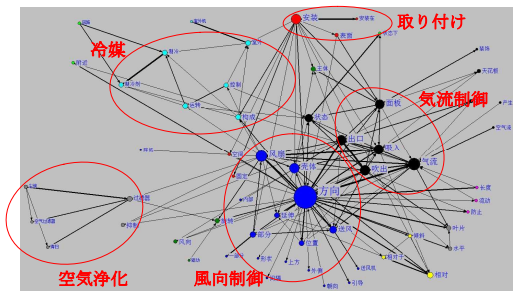


図 14. D 社の TM 結果

ここで、A 社の特徴である「风口」と D 社の特徴語「风向」に関連する KW で案件抽出を行なった (図 15)。

A 社の「风口」に関連する出願 750 件に対してテキストマイニング分析し、出現頻度の特徴語を統計分析をしたところ、「风口」に関連する出願中、「吹き出し口構造」の内容が多く、「风向制御」に関連する KW は少なかった (図 16)。

空調分野でのキーワード検索	
A 社特許：2665 件	D 社特許：588 件
KW：「风向」	
A 社：16 件	D 社：39 件
KW：「风口」	
A 社：750 件	D 社：47 件

図 15. キーワードによる検索

キーワード	頻度	キーワード	頻度	キーワード	頻度
出风口	1721	结构	392	蒸发器	243
壳体	1152	连接	389	提高	230
第二	834	位置	341	回风	229
面板	755	转动	323	内部	222
室内	735	安装	323	固定	188
部件	660	位置	315	门板	182
驱动	606	之间	304	控制	180
进风口	585	风叶	285	系统	178
风道	578	送风	278	换热	177
导风板	552	方向	278	外罩	166
绍性	541	形成	277	旋转	164
风柜	491	换热器	275	槽干	162
风口	479	侧面	265	机罩	158
结构	423	运动	253	处理	144
流风	405	电机	244	风管	140

図 16. A 社「风口」関連出願の特徴語統計

結果、中国国内メーカーの出願は、「吹き出し口構造」に関連するものが多く、逆に「风向制御」に関する技術特徴語は、日系メーカーに多く見られた。一方で、「风向制御分野」での開発は日系メーカーの方が注力してい

ると推定できると考える。

#### 4. 結論

中国語特許情報を一気通貫でテキストマイニング、統計分析、可視化できる専門のツールは現在まだないが、現存するツール及びその組み合わせである程度の分析はできることが判明した。

現在は共起ネットワーク程度しか可視化できていないが、他の可視化ツールとの連動が実現し、今後本件のような分析がより活発になることを期待する。

#### [謝辞]

本報告は「アジア特許情報研究会」でのワーキングの一環であり、会員の方々には貴重なご助言、ご協力を頂きました。ここに改めて感謝申し上げます。

#### 5. 参考文献

- [1]MTMineR.  
<http://mj.in.doshisha.ac.jp/MTMineR/> accessed 2015.10.05
- [2]ICTCLAS.  
<http://ictclas.nlp.ir.org/>  
accessed 2015.10.05
- [3]Rostnat.  
<http://weibo.com/3172254514>  
accessed 2015.10.05
- [4]Tecres Patent Map.  
<http://www.seadawn.com.cn/CN/patent.aspx?IID=25> accessed 2015.10.05
- [5]IKAnalyzerNet.  
<http://www.oschina.net/p/ikanalyzer/> accessed 2015.10.05
- [6]PatAnalyzer  
[http://www.geocities.jp/patentsearch2006/INFOPRO2014CN\\_KW\\_yokou.pdf](http://www.geocities.jp/patentsearch2006/INFOPRO2014CN_KW_yokou.pdf) accessed 2015.10.05

- [7]Cytoscape  
<http://www.cytoscape.org/>  
accessed 2015.10.05
- [8]Gephi.  
<http://oss.infoscience.co.jp/gephi/gephi.org/> accessed 2015.10.05
- [9]Ucinet6.  
<http://www.analytictech.com/>  
accessed 2015.10.05
- [10]平成 26 年度 特許出願技術動向  
調査報告書  
[http://www.jpo.go.jp/shiryuu/pdf/gidou-houkoku/26\\_8.pdf](http://www.jpo.go.jp/shiryuu/pdf/gidou-houkoku/26_8.pdf) accessed  
2015.10.05
- [11] PatentSQUARE.  
<http://panasonic.biz/it/sol/patent/>  
accessed 2015.10.05
- [12]「家用電気」、ISSN1002-5626、  
P94、空調各級別品牌格局対比（2015  
年 6 月）