

## **精度を重視した効率的な特許調査方法： 引用情報と公報の類似度に着目した特許調査方法**

○安藤俊幸<sup>1)</sup>, 平川雅彦<sup>2)</sup>  
花王株式会社<sup>1)</sup>, JFE テクノリサーチ株式会社<sup>2)</sup>  
〒131-8501 東京都墨田区文花 2-1-3  
Tel: 03-5630-9538 FAX: 03-5630-9712  
E-mail: ando.t@kao.co.jp

## **Precise and effective patent research method: Patent research method focusing on citation information and similar patents**

ANDO Toshiyuki<sup>1)</sup>, HIRAKAWA Masahiko<sup>2)</sup>  
Kao Corporation<sup>1)</sup>, JFE Techno-research Corporation<sup>2)</sup>  
2-1-3, Bunka, Sumida-ku, Tokyo 131-8501 Japan  
Phone: +81-3-5630-9538 Fax: +81-3-5630-9712  
E-mail: ando.t@kao.co.jp

### **【発表概要】**

特許調査において引用情報と特許公報の類似度を組み合わせて利用して調査精度の向上と調査の効率化を検討した。

最初に予め登録になった注目特許を選び商用データベースで引用マップと類似検索機能を検討した。①引用マップ(審査官引用)の集合と②類似検索の集合を比較すると類似検索の集合に含まれる引用特許は意外に少ないことが判明した。引用文献は審査の過程で発明のポイントが近いとして実際に使われた公報である。ただ引用情報(審査官引用)は審査が進む過程で発生するためかなりのタイムラグがある情報であり公報発行直後の段階ではあまり使えない。また審査されない公報には引用情報は無い。この引用情報の無い場合でも公報間の類似度を計算し類似の公報からスクリーニングすることによる精度向上と効率化を検討した。

引用情報と公報の類似度の応用として公報の位置関係を3次元空間上にプロットしそこに引用/被引用関係をライン表示させて同時に可視化するマップ作製方法を工夫した。

### **【キーワード】**

特許情報解析, 審査官引用, 引用マップ, 類似検索, 概念検索, 形態素解析, テキストマイニング, ネットワーク分析

## 1. はじめに

最近の企業活動のグローバル化に伴い世界の特許出願も増加している。図1にファミリー中の最初の優先日が2012年の世界の特許出願のグラフを示す。発明をファミリー単位にまとめたデータベース(Questel社 Orbit.com)のデータを使用して最初の優先国の上位10ヶ国を集計したものである。トータル1942693件のファミリーの内56%が中国(CN)で出願されている。以下同様に日本(JP)12%、米国(US)8%、韓国(KR)7%、ドイツ(DE)3%である。発明のファミリー単位とはCNの実用新案1件でも世界の数十か国に出願されていても1ファミリーになり得る可能性に注意が必要である。

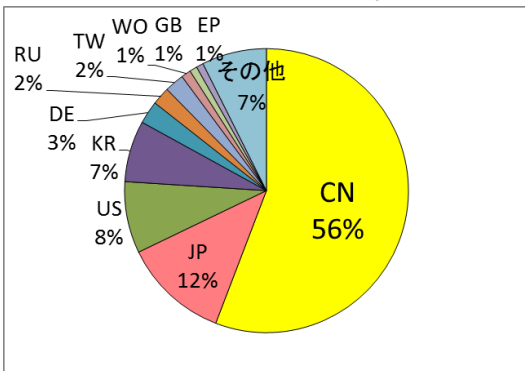


図1. 世界の特許文献(実用新案含む)

このような状況で中国特許調査の重要性は増している。また精度が良く効率的な調査方法が求められている。

昨年のINFOPRO2014では「特許価値評価の現状と課題」と題してトーク&トークが開催され、最近、経営層や知財戦略立案者向けの企業の知財面の価値を評価分析する手法が注目されている。この分野では引用情報が重視されている。また特定技術分野における出願人間の引用/被引用を解析することでその分野における競合各社の技術動向を把握したり引用的に影響が大きい重要特許を抽出できる。

INFOPRO2012-2014においてテキス

トマイニングにより中国語/日本語キーワード抽出を行い特許調査に応用した事例を紹介した<sup>1)</sup>。今回、引用情報と組み合わせで調査の効率化を検討した。

## 2. 目的

特許調査においては必要な特許の検索漏れを防ぐために網羅性を重視するとおのずと不必要なノイズが増加する。そこで目的の公報を精度よく効率的に抽出するために公報間の類似度を利用したフィルター(絞込み)やソート機能を検討する。引用情報が得られるものは利用する。

なるべく汎用性があり誰でも出来るような方法を目指す。

## 3. 方法

最初に商用データベースの引用マップと類似検索の機能概要を把握して参考にした。特定の注目特許を設定し実際の使用場面を想定して検討を行った。注目特許は特許3872179(車両の衝突防止装置)として商用データベースで引用マップ、類似検索を行い検索結果を解析した。

商用データベースは下記を検討した。NRIサイバーパテントデスク2、Questel社 Orbit.com、発明通信社 HYPAT-i2、日本パテントデータサービス JP-NET(サイテーションマップ)。

公報間の類似度情報を距離に変換して一次元上(一軸)の線形マップ、2次元平面上および3次元空間上のマップを検討した。2次元、3次元座標はR言語で非計量多次元尺度法を用いて算出してプロットした。

公報をノード、引用/被引用をエッジとした引用ネットワークマップと特定の類似度以上の公報を繋いだ類似度ネットワークマップを検討した。

(1) 解析ツール

解析には下記ツール類を使用した。

- PatAnalyzer 中国語/日本語解析ツール(自作)
- MeCab: 日本語形態素解析器<sup>2)</sup>
- saezuri lite(自然言語処理支援ライブラリ<sup>3)</sup>)
- IKAnalyzerNet: 中国語分詞ライブラリ<sup>4)</sup>
- SimCalc1 類似度計算プログラム(自作)
- R 言語: 統計解析<sup>5)</sup>
- Cytoscape: ネットワーク分析<sup>6)</sup>

解析処理の概要を図2に示す。

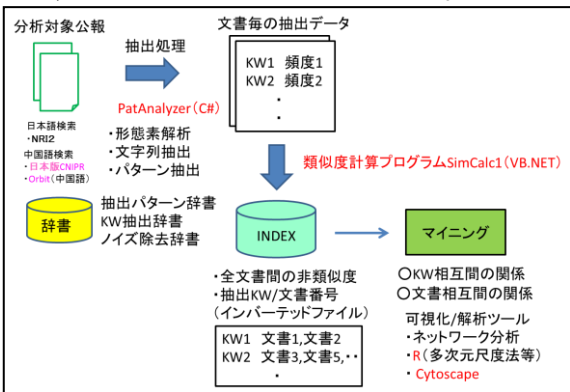


図2. 解析処理の概要

(2) PatAnalyzer 中国語/日本語解析ツール



図3. PatAnalyzer 中国語/日本語解析ツール

プログラミング言語 C#で自作したキーワード解析ツール PatAnalyzer の画面コピーを図3に示す。中国語キーワード抽

出には IKAnalyzerNet を利用している。日本語処理は saezuri lite (自然言語処理支援ライブラリ) を介して利用する形態素解析 (MeCab 和布蕪)、係り受け解析 (Cabocha 南瓜) の機能を追加している。形態素の抽出と MeCab の品詞と隣接頻度情報より求めた専門用語の抽出ができる。排除辞書に登録することで不要語のノイズ除去が可能である。

4. 結果

注目特許として特許 3872179 (車両の衝突防止装置) を設定して検討した。

4-1. 商用データベースの引用マップ

図4に NRI サイバーパテント 2 の引用マップ (一部抜粋) を示す。

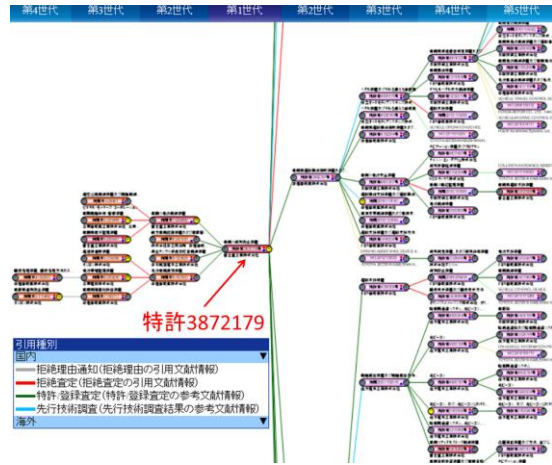


図4. NRI の引用マップ

注目特許に対して引用4世代/被引用5世代の最大 9 世代まで表示させることができる。引用/被引用関係を示す公報間の線の色が下記の状態を表している。

- 拒絶理由通知 (拒絶理由の引用文献情報)
- 拒絶査定 (拒絶査定の引用文献情報)
- 特許/登録査定 (特許/登録査定の参考文献情報)
- 先行技術調査 (先行技術調査結果の参考文献情報)

拒絶理由通知や拒絶査定等、特許/

登録査定の参考文献情報以外も表示されるのが特長である。

#### 4-2. 商用データベースの類似検索

注目特許(本願)の類似検索を各データベースで行い図5の引用特許4件の類似検索における順位とスコアを調べた結果を表1～3に示す。

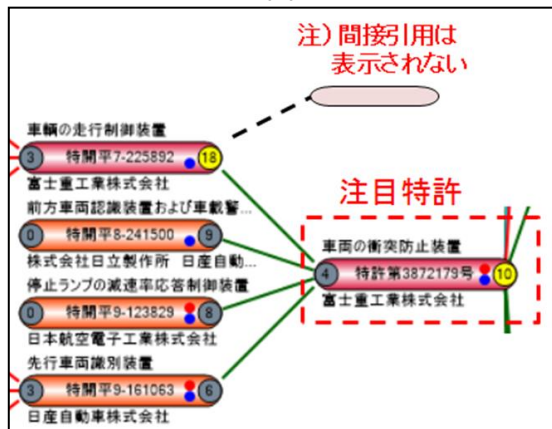


図5. 注目特許の引用と間接引用

#### NRIサイバーパテントデスク2

文献	文献番号	順位	スコア
本願	特開平11-39597	—	100
引用文献1	特開平7-225892	1145	85.2
引用文献2	特開平8-241500	7359	79.7
引用文献3	特開平9-123829	圏外	—
引用文献4	特開平9-161063	235	88.1

10000件調査

表1. NRI の類似検索結果

#### Orbit Ver.1.9.4

XPN	順位	Relevance score
JP11039597	—	—
JP07225892	38	100%
JP08241500	30	100%
JP09123829	圏外	
JP09161063	3048	80%

19999件調査

表2. Orbit の類似検索結果

#### HYPAT-i2 抄録と請求項1,対象:全文

文献	文献番号	順位
本願	特開平11-39597	—
引用文献1	特開平7-225892	78
引用文献2	特開平8-241500	—
引用文献3	特開平9-123829	—
引用文献4	特開平9-161063	—

300件調査

表3. HYPAT-i2 の類似検索結果

類似検索結果にはデータベースの特徴が表れている。引用文献3ほどのデータベースでも類似検索調査圏外である。Orbit は類似検索に特許分類を使用していると予想される。

#### 4-3. 引用/被引用と類似検索

NRI の引用/被引用マップは最大9世代表示可能であり本願を注目特許とすると国内特許192件である。

ここで図5に示す間接引用は注目特許の引用特許から被引用方向に1世代、被引用特許からは引用方向に1世代を加えた合計である。図5の間接引用は説明のために加えたもので実際には表示されない。

国内特許に注目して間接引用と注目特許の類似検索を求めると下記の件数になる。

- ①注目特許の引用マップ 国内192件
- ②間接引用1世代(①を含む) 1029件
- ③注目特許の類似検索(請求項) 上位1000件

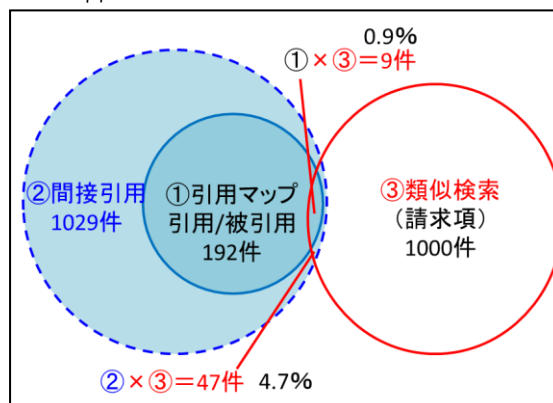


図6. 引用/被引用と類似検索の関係

図示したのが図 6 である。注目特許を様々な分野から選択し直したり、検索条件を変えても類似検索結果に含まれる引用/被引用特許は予想外に少ない。図 6 の例では類似検索結果上位 1000 件に含まれる引用/被引用 9 世代 192 件の内 9 件である。類似検索結果を分母にとると 0.9% である。間接引用 1029 件まで広げても 4.7% である。逆に 1 件の審査で引用される引用 1 世代を考えると非常に少なくなる。この例では 4 件引用されていてそのうちの 1 件が類似検索の 235 位に現れる。0.1% である。ちなみにあとの 2 件は 1145 位と 7359 位に、最後の 1 件は類似検索上位 1 万件の圏外である。

Orbit は引用文献 1、2 がそれぞれ 38 位、30 位に現れる。

審査の過程で引用される特許件数は国や分野によっても傾向は異なる。また類似検索と比べてそれほど多い数ではないという性質を踏まえても一般論としてこれからの特許検索システムには内部の仕組みは問わずに類似検索(概念検索)結果の上位に引用特許が含まれることが特許調査の効率からも望ましい。

#### 4-4. 類似度による公報関係の可視化と引用/被引用

図 7 の左側に NRI サイバーパテントデスク 2 の引用/被引用マップの全体図を示す。中心の注目特許の左側が引用特許で右側が被引用特許である。右側は国内の 192 件についてタイトル、要約、請求項までのテキストデータより相互の類似度を計算して 3 次元表示したものである。類似度が大きい公報同士が近い距離にプロットされるように 3 次元空間上の位置を多次元尺度法で求めている。引用を赤ライン、被引用を青ラインにして、注目特許と直接引用/被引用関係にある第 2 世代を太いラインで示している。類似度はテキストマイニングや類似検索で

よく使われているコサイン類似度を使用した。ブラウザで 3 次元コンピュータグラフィックスを表示させるための標準仕様である WebGL<sup>7)</sup>で出力するとブラウザを用いて 3 次元表示マップを閲覧できる。3 次元表示された引用/被引用マップをマウスで回転させて異なる視点から眺めることができる。

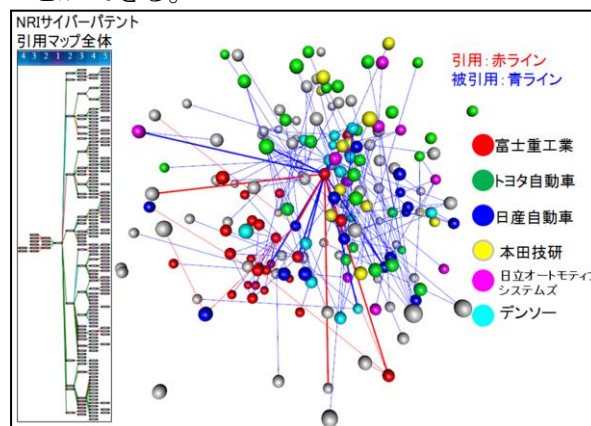


図 7. 類似度による公報関係の可視化と引用/被引用

左側の NRI の引用/被引用マップは審査の過程で発明のポイントが近いとして実際に使われた公報である。図右側で先行技術調査に類似度を使用して近い公報から確認する目的にはラインで結ばれた公報同士が近くなるように類似度の計算方法に更なる工夫が必要である。

#### 4-5. 公報の引用ネットワーク表示

図 8. は図 7 と同じ集合の Cytoscape による引用/被引用のネットワーク表示である。左側が全体図で右側は拡大図である。

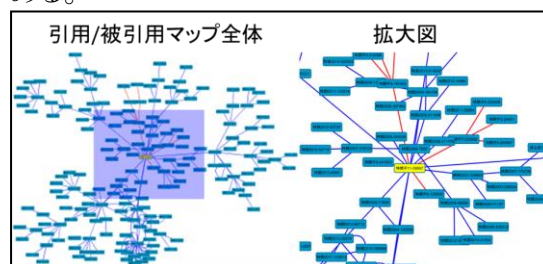


図 8. Cytoscape による引用/被引用表示

ネットワーク的な「重要度」を表す指標として次数中心性や媒介中心性等各種の中心性の指標がある<sup>8)</sup>。次数中心性(Degree Centrality)は、ある点(ノード)に接している辺の数である。媒介中心性(Betweenness Centrality)は、ある点がその他の2点を結ぶ最短経路である度合であり、値が大きいほど中心性が高い指標である。中心性は重要語(特徴語)を抽出する際の指標となる。Cytoscape ではこれらの中心性の指標に応じた可視化<sup>1)</sup>も可能である。

#### 4-6. 公報の類似度ネットワーク表示

公報の類似度を使用してネットワーク表示が可能である<sup>1)</sup>。注目公報の近くにある公報から確認することで精度重視の調査が可能になる。精度向上に向けて特徴語抽出や用語の重み付け、ネットワーク分析による公報の注目特許抽出等をさらに検討する。

#### 5. 考察

本稿では類似度計算方法の最適化検討を行っていない。類似度計算方法の検討を行い洗練させることでさらなる精度向上の余地は大きいと考えられる。

#### 6. 結論

本検討は引用情報と特許公報間の類似度を組み合わせて利用して調査精度の向上と調査の効率化を図ろうとするものである。間接引用や類似検索で調査範囲を拡大し網羅性向上を図り、類似度、引用の種類(文献のカテゴリリー等)、ネットワークの中心性の指標等でフィルター(絞込み)やソートにより調査の精度向上、効率化が可能である。

#### 7. おわりに

本稿では日本語による検討を行った。今後、中国特許調査への応用検討と類

似度計算方法の最適化検討を行い調査手法をさらに洗練させたい。

#### 「謝辞」

最後に、本報告は2015年度の「アジア特許情報研究会」のワーキングの一環として報告するものです。研究会のメンバーの皆様には様々な協力をしていただきました。ここに改めて感謝申し上げます。

#### 8. 参考文献

- [1] 安藤 俊幸ら. “中国語キーワードによる中国特許情報解析” 第11回情報プロフェッショナルシンポジウム  
[https://www.jstage.jst.go.jp/article/info-pro/2014/0/2014\\_31/\\_pdf](https://www.jstage.jst.go.jp/article/info-pro/2014/0/2014_31/_pdf)
- [2] MeCab (和布蕪) 日本語形態素解析器  
<http://taku910.github.io/mecab/>  
accessed 2015.10.09
- [3] saezuri lite.  
<http://www.vector.co.jp/soft/winnt/prog/se495669.html> accessed 2015.10.09
- [4] IKAnalyzerNet.  
<http://www.piaoyi.org/c-sharp/IKAnalyzerNet.html> accessed 2015.10.09
- [5] R 言語  
The R Project for Statistical Computing  
<https://www.r-project.org/> accessed 2015.10.09
- [6] Cytoscape  
<http://www.cytoscape.org/> accessed 2015.10.09 accessed 2015.10.09
- [7] 杉本雅広. 明解 WebGL. リックテレコム, 2015, 312p
- [8] 金 明哲. “テキストにおけるネットワーク分析”. テキストデータの統計科学入門. 岩波書店, 2009, p. 63-72.