D-024

特許文献における医学分野の因果関係抽出

Extraction of Causal Relationships of Medicine Field in Patent Documents

石川 大介 † **石塚 英弘** † **宇陀 則彦** † **藤原 譲** ‡
Daisuke ISHIKAWA Hidehiro ISHIZUKA Norihiko UDA Yuzuru FUJIWARA

1. はじめに

2002 年 7 月に策定された知的財産戦略大綱によると，これからの我が国の産業は，付加価値の高い発明などの情報を創造することが重要であると指摘されている．そのため，本研究では発明などの高度な創造性を支援できるシステムの開発を目指す．

このシステムの開発のため，(a) 知識抽出，(b) 知識統合，(c) 知識利用，の項目別に研究を行なう．(a) として，特許文献から因果関係の抽出を試みる．繊維工学の分野を例にした場合，化学物質とその性質の関係を抽出した [1]．(b) として，新聞などの情報源との連動を図る．(c) として，抽出した因果関係を類推に利用する．抗菌性を例題とし，特許文献を利用した類推方法を提案した [2]．

本稿では，(a) として，特許における医学分野 (医薬用，歯科用，化粧用製剤) を例に，因果関係の抽出を試みる．この抽出方法と抽出結果を報告する．また，(b) の例として，抽出結果と新聞との連動について述べる．

2. 因果関係の抽出方法

2.1 対象文献

2002 年 NTCIR ワークショップの NTCIR-3[3] において，特許コーパスが配布された．本研究ではこのコーパスから，Japio 出願抄録の 1998 年度版 (日本語，34 万件) を利用する．この 34 万件から、国際特許分類 IPC コードの A61K「医薬用，歯科用，化粧用製剤」が付与された文献 5205 件を実験に使用する．

これら特許文献の概要部分を使用することとする．概要部分は完結にその発明の内容が記述されているからである．以下に文献の例を示す．

1. 資生堂：(株) , 特願 1996-188234
（J）クララ抽出物を含有する ことにより、エラスターゼ阻害作用を有し、皮膚のハリや弾力を保持し、若々しい肌の状態を維持可能とする。

2. 協和醗酵工業 (株) , 特願 1996-180223
（J）特定のジベンゾオキセピン誘導体を有効成分として用いる ことにより、優れた抗アレルギー作用を有し、気管支喘息、アレルギー性鼻炎、アトピー性皮膚炎、じん麻疹等に有効な薬剤を得る。

文献 1 からは「クララ抽出物 → エラスターゼ阻害作用」，文献 2 からは「ジベンゾオキセピン誘導体 → 抗アレルギー作用」という関係が伺える．以下，このような物質とその作用の関係を抽出する方法について述べる．

†筑波大学大学院図書館情報メディア研究科
‡独立行政法人工業所有権総合情報館

2.2 「ことにより」表現

前記のどちらの文献も「手段に関する記述部〜ことにより〜効果に関する記述部」という表現で記述されている.「ことにより」という記述について，医学分野の特許文献 5205 件を調査したところ，3041 件の文献に使用されていた．以後,「ことにより」が使用されている 3041 件を対象文献とする．なお,「ことにより」が使用されていない文献 (2164 件) は，効果のみ，あるいは手段と効果の記述が不明なため，本研究では扱わないこととする．

この対象文献 3041 件に対して，形態素解析システム茶筌 [4] を利用し,「名詞、未知語、記号」の形態素で構成されている語の集合をひとまとまりの用語と見なし，語分割テキストを生成した．

2.3 物質名の抽出

頻度	手段の用語
1570	特定
1254	含有
540	有効成分
345	配合
159	化合物

頻度	効果の用語
250	〜作用
220	〜効果
81	〜成物
78	〜疾患
78	〜可能

図 2: 手段と効果の記述部に出現する用語

前記の文献の例では,「ことにより」より前半が手段の記述部であり，使用された物質が記述されている．手段の記述部で物質を表わさない用語は，文献 1 では「含有」，文献 2 では「特定，有効成分」である．これら物質を表さない用語は出現頻度の高い用語と考えられる．手段の記述部に出現した用語の頻度を集計したものを図 2 に示す．

本研究では，この上位 4 位までの用語 (特定，含有，有効成分，配合) を手段の記述部から取り除き，残った用語が一つの場合にその用語を使用された物質名と考えて抽出する．なお，複数の用語が残った場合，どの用語が手段の要因になるかは判断に難しいため，本研究では対象外とした．

2.4 作用表現の抽出

前記の文献の例では,「ことにより」より後半が効果の記述部であり，発現した作用が記述されている．効果の記述部で発現した作用は，文献 1，2 ともに「〜作用」という用語で表現されている．

効果の記述部に出現した用語の末尾二文字の頻度を集計したところ，図 2 に示すように「〜作用」の他に「〜効果」という用語が高頻度で使用されていることが分かった．

本研究では，効果の記述部に現れた「〜作用」と「〜効果」の用語を，発現した作用を表現する用語 (以下，こ

手段	→	効果	抽出元の特許文献
４－ヒドロキシチオフエノール誘導体	→	美白作用	御木本製薬　（株）, 特願 1996-311168：（Ｊ）４－ヒドロキシチオフエノール誘導体を含有することにより、美白作用を向上する。
栗皮抽出物	→	美白効果	コリアナ　コスメティツクス　ＣＯ　ＬＴＤ, 特願 1998-029417:（Ｊ）栗皮抽出物を含有することにより、安全性に優れ、皮膚改善及び美白効果に有効な組成物を得る。
メトキシフエノール化合物	→	美白効果	資生堂：(株）, 特願 1997-072680：（Ｊ）特定のメトキシフエノール化合物を含有することにより、ハイドロキノン以上に美白効果を発揮し、且つ、安全性の高い外用剤を得る。
オウゴニン	→	メラニン生成抑制作用	ヤクルト本社：(株）, 特願 1997-108353：（Ｊ）オウゴニンを有効成分として含有することにより、安定性、安全性及びメラニン生成抑制作用等を向上する。

図 1: 美白に関する抽出結果の例

れを作用表現と呼ぶ) として抽出する．なお，効果の記述部に現れた作用表現は全て抽出するものとする．

2.5　物質-作用関係の抽出

手段の記述部から物質名を，効果の記述部から作用表現を抽出できれば，抽出された物質名と作用表現との関係は，因果関係を示すと考えられる．本研究では，この物質名と作用表現の関係を物質-作用関係と呼び，これを抽出する．

3.　抽出結果

対象文献 (3041 件) から物質-作用関係が 178 件抽出された．この詳細は，物質名は抽出されたものが 502 件，作用表現は抽出されたものが 889 件であった．

抽出結果から，美白に関連する例を図 1 に示す．

4.　考察

抽出結果 178 件が示す物質-作用関係は，抽出元の文献の記述通り適切かを調査した．なお，抽出された物質名および作用表現が曖昧な場合は誤りと見なした．適切と判断できた抽出結果は 81 件である．適切と判断できない抽出結果の例を以下に示す．

1. 物質名が曖昧：化合物，植物，新規化合物，など
2. 作用表現が曖昧：効果，作用，改善効果，など
3. 作用表現の否定や間接表現：副作用，など
4. その他，形態素解析のミス，など

1 は，抽出元の文献にもともと物質名が明確に記述されていないことが原因である．抽出処理は適切であった．

2 の例として,「〜の効果」といった表現があった．複数の用語からなる作用表現は今後の課題である．

3 の主な原因は作用表現「副作用」である．この作用表現は,「副作用が少ない」といった否定で表現される．今回の抽出結果を確認したところ,「副作用」は全て否定で表現されていた．その他の 3 に該当する抽出結果は 4 件あり，件数は少ないが今後の課題である．

5.　他の情報源との連携

日経全文記事データベース [5] において,「美白」を検索すると，次の記事が見つかった．

> 「資生堂、美白効果高めた美容液（情報プラス)」(産業 2002.6.21)
> (〜略〜) 新商品はシミなどの原因となるメラニンの生成を抑制するアルブチンに加え、抗酸化作用を持ち活性酸素を除去するビタミンＣエチルを新配合、美白効果を高めたという。(〜略〜)

この記事から，メラニンの生成を抑制することで，美白効果が得られることが分かる．よって，図 1 の「オウゴニン」も美白効果があると判断できる．このように，抽出結果と他の情報源を利用することで，抽出結果の新たな情報を得ることができる．

6.　おわりに

本稿では，医学分野を例に因果関係の抽出を行なった．本手法は，複雑な構文解析を必要としないことが特徴である．今後は，複数の物質が関係する手段の記述の分析を行なう．

謝辞

本研究では，国立情報学研究所で作成された NII-NACSIS コレクションの NTCIR3 を使用いたしました．深く感謝いたします．なお，本研究の一部は日本科学協会笹川科学助成金 (研究番号 16-431) からの助成を受けています．

参考文献

[1] 石川大介，石塚英弘，宇陀則彦，藤原譲：特許文献を用いた因果関係に基づく知識構造化の試み，情報処理学会研究報告 2003-FI-72, pp.61-67, 2003.

[2] 石川大介、石塚英弘、宇陀則彦、藤原譲：特許文献およびそれから抽出した因果関係を用いた類推-思考支援システムを目指して-，人工知能学会研究会資料 SIG-FAI-A301-02, pp.7-12, 2004.

[3] NTCIR：http://research.nii.ac.jp/ntcir/index-ja.html

[4] 形態素解析システム　茶筌：http://chasen.aist-nara.ac.jp/index.html.ja

[5] 日経全文記事データベース「日経産業新聞・日経金融新聞・日経 MJ（流通新聞)」，2002 年版，日本経済新聞社.