

手がかり句を用いた特許請求項の構造解析および分散表現による類似特許検索

Yaegashi, Temma / 八重樫, 天真

(出版者 / Publisher)

法政大学大学院理工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学研究科編

(巻 / Volume)

64

(開始ページ / Start Page)

1

(終了ページ / End Page)

3

(発行年 / Year)

2023-03-24

(URL)

<https://doi.org/10.15002/00026411>

手がかり句を用いた特許請求項の構造解析 および分散表現による類似特許検索

STRUCTURE ANALYSIS OF PATENT CLAIMS USING CLUE PHRASES
AND ASSOCIATIVE PATENT RETRIEVAL BY DISTRIBUTED REPRESENTATION

八重樫天真

Temma Yaegashi

指導教員 藤井章博

法政大学大学院理工学研究科応用情報工学専攻修士課程

Patent information is a valuable source of information for research into existing research and understanding the latest research trends, and it is also necessary to conduct a thorough search to avoid infringing on the patent rights of other companies. However, since there are a huge number of patents and it is difficult for non-specialists to grasp the contents of patents due to their unique descriptions, natural language processing technology is expected to be utilized. In this study, as a support for patent searches, the structure of claims is analyzed based on clue clauses, and similarity judgments are made based on the distributed representation of BERT. As a result of the experiment, an accuracy of 0.406 was obtained for the similarity evaluation with patents in the reference literature. Translated with www.DeepL.com/Translator (free version)

Key Words :PATENT, STRUCTURE ANALYSIS, BERT

1. はじめに

企業や研究者が新規開発を検討する際に、既存研究の調査や最新の研究動向の把握は重要である。一方で、研究開発に関する情報は企業にとって戦略的に重要なものであり、内部情報として秘匿される。その中で、出願公開によってその発明の内容が明らかになる特許情報は貴重な情報源であると共に、他者の特許権を侵害しないためにも綿密な調査が必要である。

特許の読解はその記述の特殊性から非専門家にとって内容の把握が困難である一方で、特許数は膨大であり、既存の特許プラットフォームでは検索精度が甘く、その精査に膨大な時間がかかってしまう。そのため、特許文書における自然言語処理技術の活用が期待されている。

本研究では、特許庁が提供する特許情報標準データに2023年1月10日に公開された446個の登録特許とその特許の参考文献に記された特許の対を用いる。特に、権利範囲を明示する特許請求項の文章を対象に構造解析による主題と技術構成の情報の抽出とそれを用いた類似性の判定を試みる。

2. 関連研究

(1) 特許検索に関する研究

自然言語処理分野では単語の意味を定量的に表現できる分散表現や文書分類、検索、翻訳など多くの領域で技術発展がなされる中、特許文書においてもその活用が期待されている。一方で、特許の記述の特殊性や、特許に記載される技術の専門性の高さによる特許文書の解析には大きな課題がある。また、他者の特許権の侵害を防ぐために綿密な調査が必要であり、先行研究では検索システムに対して、発明の現場にいる弁理士からの一定の評価は得られたものの、そのシステムで得られた特許のみでは不安であるという声が上がった。

本研究では、検索結果の根拠を示すために請求項から特許を構成する技術を抽出し、どの技術に類似した特許が存在し、どの技術が本研究システムで類似性の高い技術を検索できなかったかを示す。

(2) 請求項の構造解析に関する研究

請求項は係り受け構造や入れ子構造、並列表現や箇条書きなどの多種多様な構造を組み合わせで表現されている。そのため、通常のテキスト解析とは異なり、請求項特有の構造を利用した手法のほうが制度が良いことが報告されている。先行研究では手がかり句を用いた構造解析や系列ラベリング問題として自動でタグを付与する手

法が提案されている。

本研究では、請求項が主題、構成技術、詳細な説明の3つの要素から成り立っているととし、「～であって、」や「～を特徴とする」などの手がかり句を用いて、請求項の3要素への分割と、「～し」や「～と」などの手がかり句を用いて、構成技術の分割を行う。図1に示すように手がかり句からの構造解析を試みる。

(主題)特許請求項から特許の類似性を判断するシステム**であって、**(技術1)手がかり句をもとに請求項を主題、構成技術、詳細な説明の3要素に分割する構造解析手法**と、**(技術2)BERTの分散表現を用いたcos類似度によって、特許の類似性を判定する類似度評価手法**からなり、**(説明)上記類似度評価手法は上記構造解析手法によって、得られた主題と構成技術の2つの要素で類似性を判定することを**特徴とする**システム

図1 本研究で行う構造解析例

3. 研究手法

本研究では、以下の4つの手順により、データベースの中から類似する特許の検索を行う。

(1) 手がかり句を用いた構造解析

以下に、特許請求項を主題、構成技術、詳細な説明の3要素に分割する手順を示す。

- ① 「～であって、」「～において、」の表現が出現した場合に文章を分割する。
- ② 分割した前半の文章に「～を特徴とする」「～からなる」のような表現が出現した場合、前半の文章を構成技術とし、出現しなかった場合、前半の文章を主題とする。
- ③ 構成技術が決定した場合、後半の文章を「～を特徴とする」「～からなる」などの表現によって、再度文章を分割し、前半を詳細な説明、後半を主題とする。
- ④ 主題が決定した場合、「～を特徴とし、前記～は」などの表現で文章を分割し、前半を構成要素、後半を詳細な説明とする。

(2) 手がかり句を用いた構成技術の分割

(1)の手順によって、抽出した技術構成において、読点の前の表現「～と、～と」や「～し、～し」、「～部、～部」に着目し、構成技術の分割を行う。

(3) 分散表現による主題の類似評価

図2に示すように、主題を比較し、BERTの分散表現+cos類似度を利用し、類似度が一定以上であれば、(4)の手順を行う。

(4) 分散表現による構成技術の類似評価

(3)によって、主題の類似性が高いと判断された特許について、互いの構成技術を総当たりで類似度評価を行い、1つ以上類似度が一定以上となれば類似した特許と判

定する。

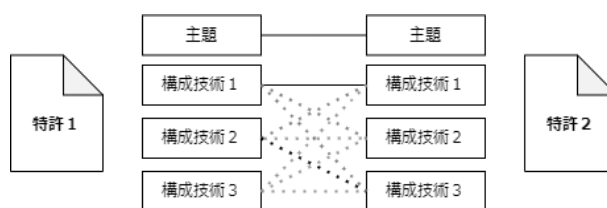


図2 類似度評価手法

4. 実験

特許庁が提供する特許情報標準データ446件の特許と、その特許の参考文献に記された特許を提案手法によって、構造解析と類似度評価を行い、対の特許の類似性を判定する。

(1) 結果

表1に提案手法によって、特許請求項の構造解析を行い、参考文献として対となる特許との類似度判定を行った結果を示す。

表1 提案手法によって類似性が高いと判定できた割合

主題	主題+構成技術
0.471(210/446)	0.406(181/446)

(2) 考察

今回類似度が低くなった対に関して、読点のない文章や、人間が見てもわかりにくい複雑な構造によって、構文解析手法がうまくいっていない場合が多かった。また、特許によって、文章の長さがまちまちであるために請求項の内容のみでは情報が不十分であると感じた。今後は請求項以外の要素も使い、一定の長さで特許の趣旨が伝わりやすい主題と、構成技術の抽出を試みたい。

また、特許検索だけでなく、背景技術などの記載から技術動向の調査など、特許分析によって得られる知見の可能性も探っていきたい。

5. 結論

本研究では、特許庁が提供する特許情報標準データとその参考文献を用いて、特許請求項の構造解析および、類似度算出を行った。実験の結果、0.406の精度が得られた。

謝辞：本研究において様々なご指導およびご協力いただいた応用ネットワーク研究室の藤井章博教授、藤井研究室内の皆さんにとっても感謝いたしております。

参考文献

- 1) 新森昭宏, 奥村学, 丸山雄三, 岩山真. 手がかり句を用いた特許請求項の構造解析, 情報処理学会論文誌, 2004, vol. 45, no. 3, p. 891-905

- 2) 白坂一. AI と弁理士の協働による進歩性判断—先行特許文献調査システムの発明現場への導入—」北陸先端科学技術大学院大学先端科学技術研究科 博士論文 2021
- 3) 難波英嗣. 類似内容の特許請求項の自動対応付け, 情報処理学会, 研究報告情報基礎とアクセス技術 (IFAT), 2021, vol 142, no 7, p1-4