

1-14 IP ランドスケープ支援のための特許情報のベクトル化を用いた共起語ネットワーク作成システム

奥原研究室研究室
2020032 平井 遥斗

1. はじめに

近年、経営環境は大きく変化しており、いわゆるVUCAな時代を迎えており、企業が持続的な発展を図るために、自社の核となる独自の強みを生かし、他者との差別化を図ることが極めて重要である。そんな中、IPランドスケープが注目を集めている。[1]。

本研究では、今日に至るまでの莫大な特許文章群を対象とした知見発見および探索を目的とする。

2. 知的財産戦略と特許情報処理

知的財産戦略は、経営戦略と密接に関係しており、企業全体の戦略において各部門や機能の方向性を決定する重要な役割を果たしている。日本において、知的財産戦略は特許などの知的財産「Intellectual Property(IP)」と景観や風景を意味する「Landscape」を組み合わせた造語で「IP ランドスケープ」と呼ばれることが多い。特許庁の調査によれば業界における IP ランドスケープの必要性は大半が認知しているものの十分に実施できている企業は少ない[2]。実践面でのハードルが高いと言わざるを得ない。

特許情報処理とは、特許文章や特許データベースの情報をコンピュータで自動的に解析・処理を行うことで、莫大な量の特許情報から有用な知見を効率的に見出し、知的財産戦略の立案や技術動向の分析などに活用することを目的としている。

3. 特許情報の可視化

莫大な量の特許情報をベクトル化し、可視化することで特許の技術分野や技術トレンドなどを把握することができると言える。具体的には特許本文を Sentence-Bidirectional Encoder Representation from Transform (Sentence-BERT) を用いることで文章全体を単位にベクトル化を行う。また、次元の呪いを回避するために次元圧縮を用いたのち、クラスタリングによって潜在的なグループ化を行う。

次元圧縮手法には、線形次元圧縮手法と、非線形次元圧縮手法がある、本研究では、ベクトル同士の近さを保有する必要があるため非線形圧縮手法を用いる。非線形圧縮手法には Uniform Manifold Approximation and Projection of Dimension Reduction(UMAP) を用いる。

関連性の高い単語は、一緒に出現することが多いた

め、それらの単語の共起関係を調べることで、単語間の関係性を理解することができる。具体的には共起関係にある単語と単語を線で結んで描かれる共起語ネットワークを作成する。

4. 提案手法

ユーザーが入力したキーワードをもとに Google Patents での検索結果をスクレイピングし、それらのデータに対してベクトル化やクラスタリングの処理などを行う。またそれぞれのクラスターに対してクラスターの解釈に値するタイトルを与える。ユーザーはそれと散布図を照らし合わせることで自分が見たいクラスターを選択することができる。

選択されたクラスターにおける共起語ネットワークを Simpson 係数を求ることで作成する。作成された共起語ネットワークは Three.js および pyvis を用いて 3D グラフと 2D グラフによって可視化を行う。

5. シミュレーション結果ならびに考察

「ブロックチェーン技術を持っている会社の、技術の決済システムへの使い道の検討と現状の把握」という事例を題材に提案システムを使用した。

実際に作成された散布図と共に起語ネットワークを図4に示す。

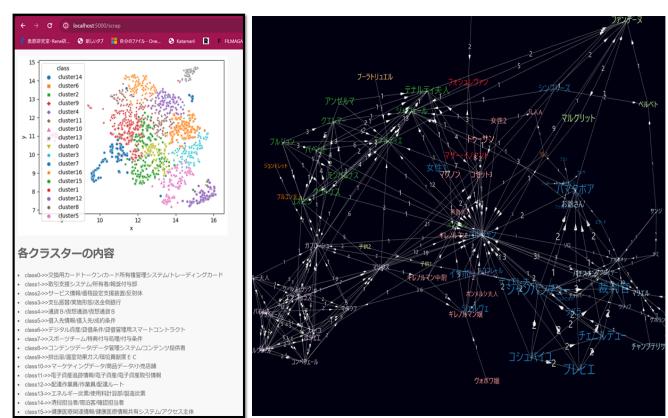


図4 キーワード「ブロックチェーン」,「決済システム」での出力結果

6. おわりに

本研究では、特許文章に対してテキストマイニングやトピックモデリングといった自然言語処理技術を適用することで、IP ランドスケープ実施への支援を行った。今後の課題として、システムの処理時間の短縮や

参考文献

- [1] <https://www.jpo.go.jp/news/koho/kohoshi/>.
- [2] <https://www.jpo.go.jp/support/general/document/chizaijobobunseki-report/chizai-jobobunseki-report.pdf>.