進捗報告

平井　遥斗

富山県立大学　情報システム工学科

2023 年 10 月 20 日

はじめに

IP ランドスケープ

IP ランドスケープ（Intellectual Property Landscape）とは知的財産情報を分析しその結果から経営戦略の策定や企業の意思決定に活用することのほか，知的財産を重視した経営を指す.

現状

2021 年に特許庁が公表した「経営戦略に資する知財情報分析・活用に関する調査研究」では国内企業等 1,515 者に対してアンケート調査を行っている.
IP ランドスケープを実施し，その結果，経営層等に共有できていると答えた人は 10%程であった.
一方で，約 80%が IP ランドスケープは必要と答えており,「必要性はわかっているがなかなか実施に至らない」という企業が多い実態がわかる.

特許情報のベクトル化

特許情報

GooglePatants から特許情報をスクレイピングした.
GooglePatents は最大 1000 件までしか検索結果を出力できないので，年代ごとにわけでスクレイピングを行った.
「歯ブラシ」と「歯磨き粉」をキーワードにしたところ，約 1 万個の特許を取得することができた.

BERT

従来の自然言語処理技術よりも優れており，文脈から意味を理解することができる.
それぞれの文章を 768 次元のベクトルとして表現することが可能.

次元圧縮とクラスタリング

umap を用いて 15 次元まで次元を圧縮したのち，k-means を用いて 15 のクラスターに分けた．

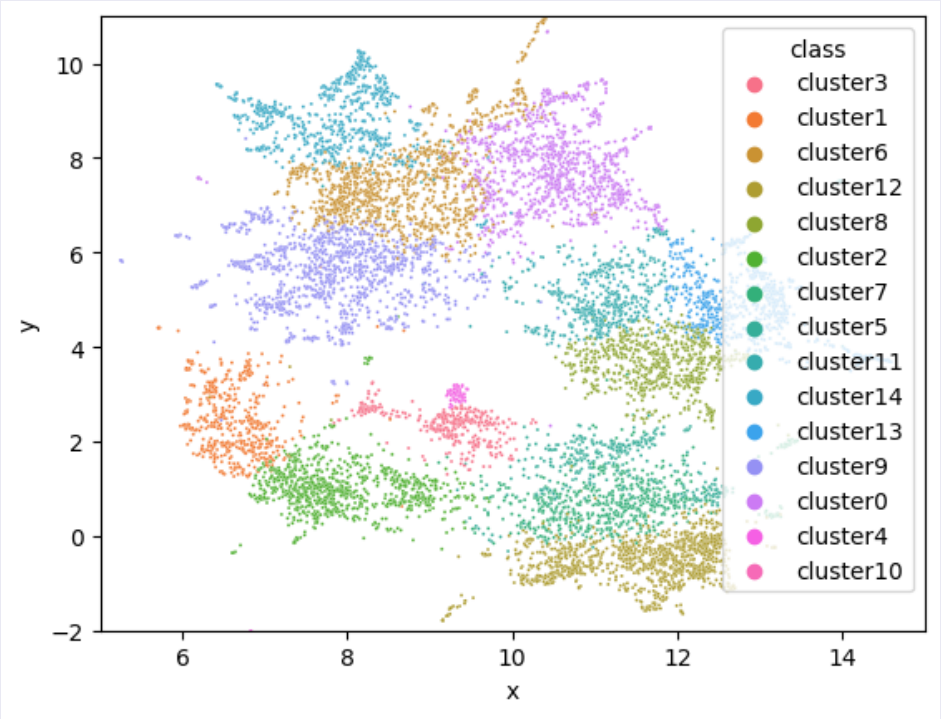

図 1: ベクトル (2 次元)

クラスターの数がこれで正しいのかわからないので，X-means や G-means を使うことも視野に入れている．

クラスタリングと可視化

クラスターのタイトル

前回の発表でクラスターのタイトルがあった方がいいのではという指摘があったのでタイトルを作成した.
tf-idf を用いてクラスター内の文章の中から重要な単語を上位 3 語表示した.

図 2: クラスターのタイトル

共起語ネットワーク

単語間のつながり

単語の共起関係

- Python は，機械学習でよく利用されるプログラム言語である．
- Python は，機械学習だけではなく，Web アプリ開発などでも利用されている．

この時「Python」と「機械学習」は共起関係にあるといえる．
この共起関係の強さを Jaccard 係数を用いて導出した．

共起語ネットワーク

共起語ネットワーク

Jaccard 係数をもとに共起語ネットワークを作成し 3D グラフで表示した.

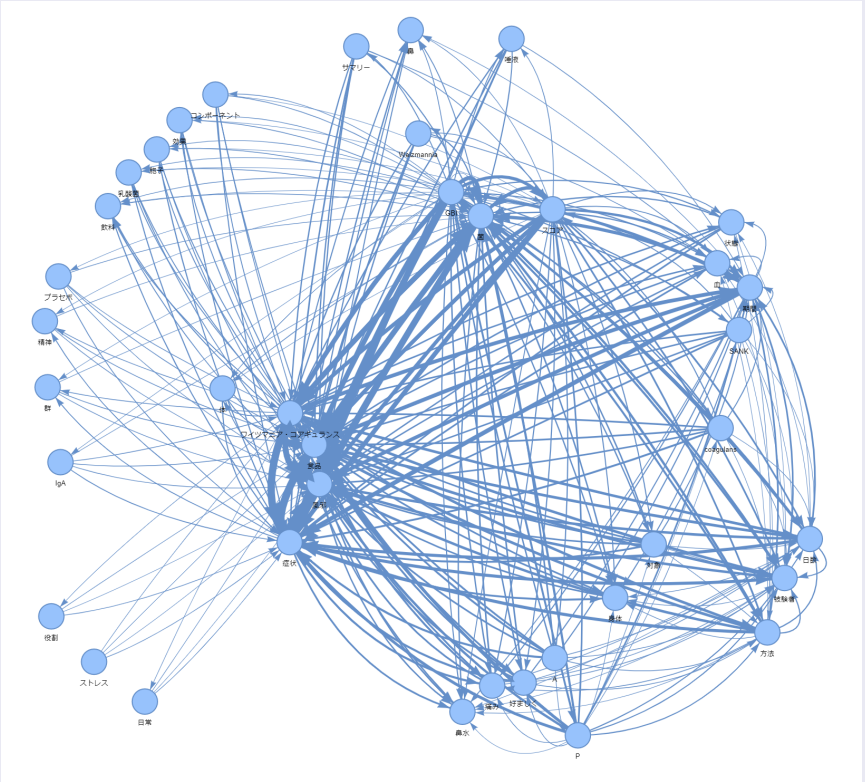

図 3: 共起語ネットワーク（可視化)

まとめ

まとめ

BERT や umap を勉強しているが難しいところが多く理解できていない部分があるので，質問されたときにちゃんと答えられるように，さらに理解度を高めたい．
特許には専門用語が多く，現在の辞書では，ちゃんと拾えていない単語があるため，専門用語の抽出を行ったが，まだ辞書に登録するところができていないので，これから実装したい．