

はじめに

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

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。

IP ランドスケープの活用

3/11

IP ランドスケープの活用

すでにある技術を、さらに有効活用できるマーケットを探る。

→会社が持っている技術の活用先を特許分析を行うことで見出すことができる。

足下の課題解決につながるような研究開発テーマを探る。

→会社が抱えている問題を解決できる技術を特許分析によって発見することができる。

ライセンス先・提携先の候補を探る。

特許情報のベクトル化

4/11

はじめに

特許情報

GooglePatents 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、年代ごとにわけて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。

クラスタリングと可視化

5/11

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。

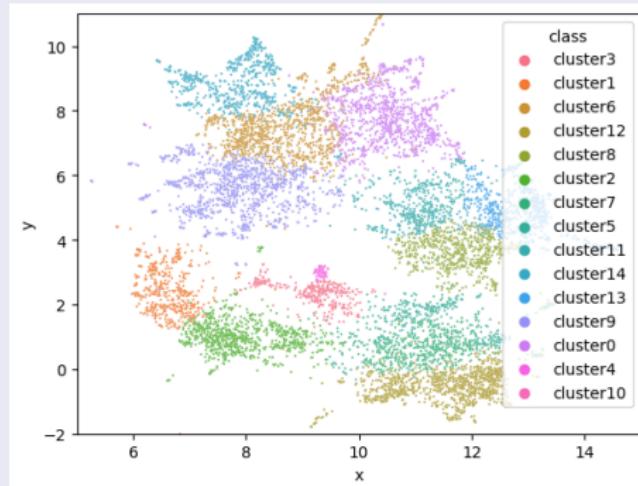


図 1: ベクトル (2 次元)

クラスターの数がこれで正しいのかわからないので、クラスター分析を行った。

はじめに

シルエット分析

シルエット係数

凝集度：データ $x^{(i)}$ が属するクラスタ C_{in} の他の点までの平均距離

$$a^{(i)} = \frac{1}{|C_{in}|-1} \sum_{x^{(j)} \in C_{in}} \|x^{(i)} - x^{(j)}\| \quad (1)$$

乖離度：データ $x^{(i)}$ に最も近い別クラスタ C_{near} に属する点までの平均距離

$$b^{(i)} = \frac{1}{|C_{near}|} \sum_{x^{(j)} \in C_{near}} \|x^{(i)} - x^{(j)}\| \quad (2)$$

シルエット係数

$$s^{(i)} = \frac{b^{(i)} - a^{(i)}}{\max(a^{(i)}, b^{(i)})} \quad (3)$$

クラスタリング分析

7/11

シルエット分析

シルエット分析結果

はじめに

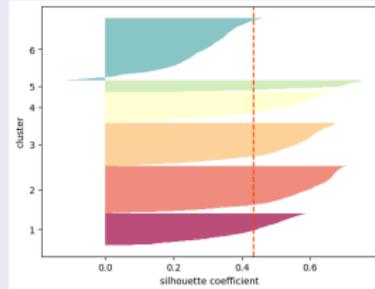


図 2: 6 クラスター

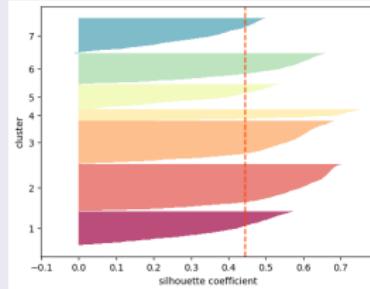


図 3: 7 クラスター

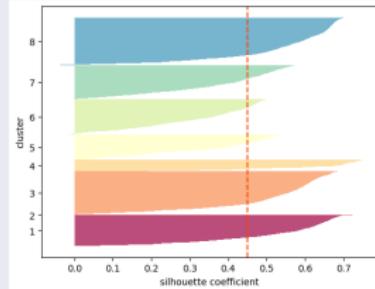


図 4: 8 クラスター

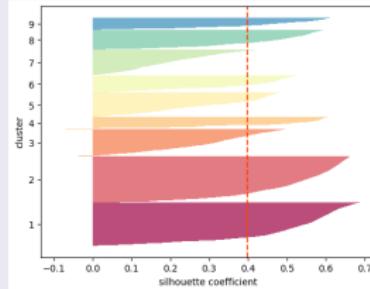


図 5: 9 クラスター

クラスタリングと可視化

8/11

クラスタリング

シルエット分析の結果から、7，8 クラスが妥当であることが分かった。

このことから、8 つのクラスターに分けた。

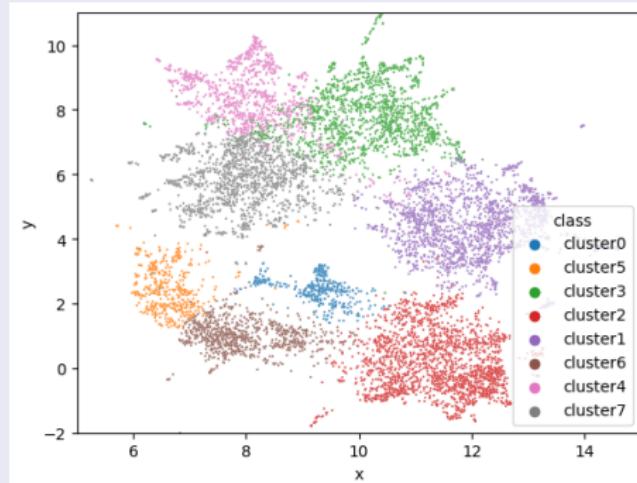


図 6: クラスター (8 つ)

はじめに

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。

他にも共起関係を導出する係数として、Dice 係数や Simpson 係数があり、それらを用いた場合も検証する必要がある。

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。

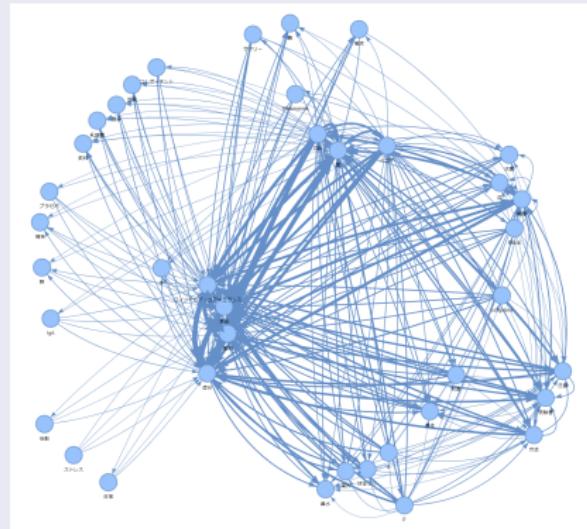


図 7: 共起語ネットワーク (可視化)

まとめ

11/11

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。