

自己組織化マップによる特許マップの作成

河野 久志[†] 小原 和博[‡]千葉工業大学大学院工学研究科^{†‡}

1. はじめに

特許マップとは、特許情報を特定の利用目的に応じて収集、分析し、図面、グラフ、表などで視覚的に表現したものである。多数の特許を対象に、テキストマイニング[1]を用いて、課題と解決手段の 2 軸マップ[2, 3]や、時間とビジネスインパクトの 2 軸マップ[4]を作成した研究報告がある。我々はテキストマイニングと自己組織化マップ (SOM) [5]を用いて、企業毎の特許マップと、特許毎の特許マップを作成した。2 軸マップでは縦軸、横軸に明確な意味があるが、SOM では縦軸、横軸には意味がない。SOM ではトータルで見て類似の企業や特許が近くに配置されるので、企業間あるいは特許間の関係を視覚的に調べるのに適していると考えられる。

2. 特徴ベクトルの抽出

特許の特徴ベクトル抽出法として次の 3 つを考えた。①単語頻度解析の結果を利用し、頻度上位の単語を抽出 (但し、似た意味合いの単語を一つにまとめて次元数を削減)。②手がかり語を用いて特徴語を抽出。③上記①と②の組合せ。いずれも、該当する単語数を属性値とする。

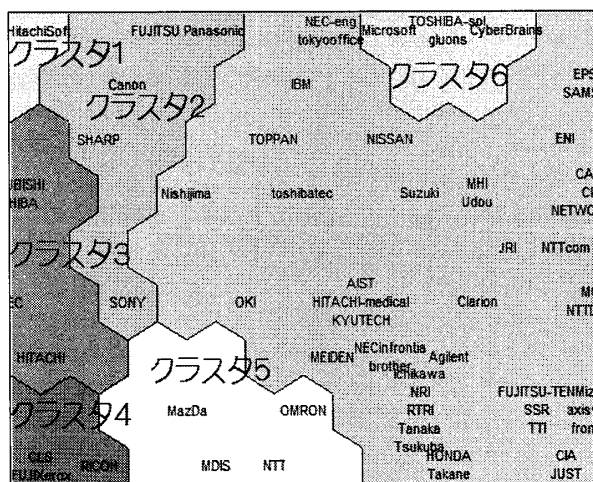


図 1 単語頻度解析結果を用いた特許マップ

Creating Patent Maps with Self-Organizing Maps

[†]Hisashi Kouno, [‡]Kazuhiro Kohara^{†‡}Graduate School of Engineering, Chiba Institute of Technology

2.1 単語頻度解析を利用

Text Mining Studio を用いて、SOM 作成に用いる特徴ベクトルを抽出した。使用した特許公報は、特許電子図書館で検索した「文章解析」と「テキストマイニング」に関する 217 件。出願人の数は 67。分析対象としたのは「要約の課題」と「要約の解決手段」である。まず単語頻度解析を行い、頻度 5 以上の 46 語を抽出した。「コンピュータ」や「対象文」「1 つ以上」などの不適当な語を削除し、似た意味合いの単語を一つにまとめた結果、13 語になった単語を特徴ベクトルの属性とした。例えば「文章」「音声」「ネットワーク」「データベース」である。

2.2 手がかり語から特徴語を抽出

次に、構文解析を利用した分析として、手がかり語に係る単語から特徴語を抽出した。手がかり語としたのは「本発明」「提供」「課題」「目的」の 4 語である[2]。これらに係る単語から、特徴語にふさわしい単語 20 語を選出した。例えば「電子番組ガイド情報」「電子ブックプレーヤ」「記事配信装置」である。

3. 企業毎の特許マップの作成と分析

企業毎に全特許の特徴ベクトルを加算してから、Viscovery SOMine 4.0 を用いてマップを作成した。ノード数は 100 とした。第一に、単語頻度解析の結果を入力 (13 次元) として特許マップを作成した (図 1)。6 つのクラスターに分かれている。属性毎の分類マップである「属性マップ」を見ると、クラスター 1 では「リスト」という単語が多く使われており、クラスター 2 では「文章」「音声」といった単語が、クラスター 3 では「テキストマイニング」「ユーザ」などの単語が使われている。クラスター 4 では「高精度」、クラスター 5 では「組み合わせ」、クラスター 6 では「効果的」という単語が使われている。

第二に、手がかり語から抽出した特徴語を入力 (20 次元) として特許マップを作成したが、特徴ベクトルがスパースなため (特徴語が使われている特許が少ないため)、まばらなマップとなってしまった。

第三に、頻度単語 13 語と特徴語 20 語を組み合わせ、33 語の特徴ベクトルを入力 (33 次元) としてマップを作成した (図 2)。

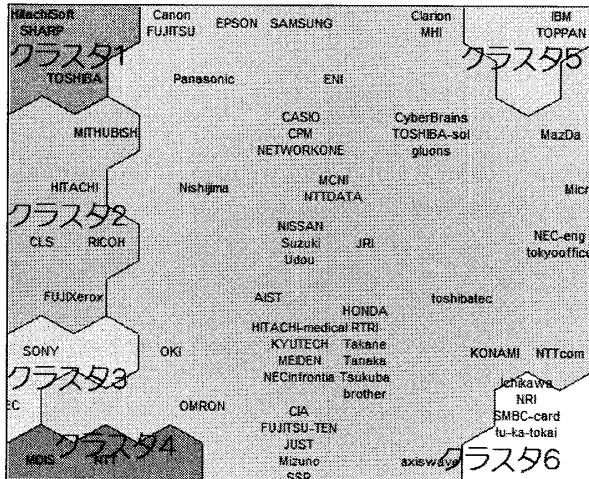


図 2 単語頻度解析と特徴語による特許マップ

図 1 と同様、6 つのクラスタに分かれている。このマップから、クラスタ 1 の企業では「ユーザ」の嗜好を反映した「電子番組ガイド情報」を登録する記録媒体、クラスタ 3 の企業では文字テキストを「音声」として出力する「電子ブックプレーヤ」や「テキストマイニング」を利用して記事を分類する「記事配信装置」、クラスタ 5 の企業では「データベース」から検索条件にあった「マーケティング」データを提供するプログラムなどを作っていることがわかった。頻度上位の単語と特徴語を組み合わせたことにより、企業の分類と各クラスタの具体的な傾向を知ることができた。

また、特定の企業における、マップ上での近傍ノードを表示した（図 3）。

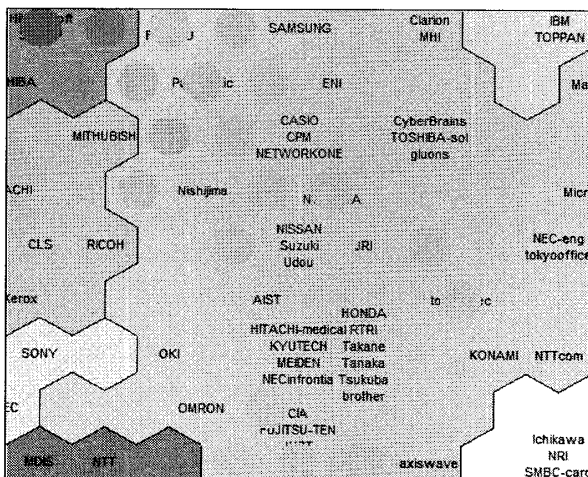


図 3 S 社の近傍表示

左上の濃い赤丸が S 社である。E 社、P 社、M 社、N 社などと「ネットワーク」の属性で近い関係にあることがわかる。近傍表示により、企業間の関係をピンポイントで調べ、特定企業の特許戦略に活用できると期待する。

4. 特許毎のマップの作成

単語頻度解析と特徴語を組み合わせた 33 語を使って、特許毎の関係を示す特許マップを作成した（図 4）。各特許を対象にすることで、特許全体の動向を視覚的に調査できると期待する。

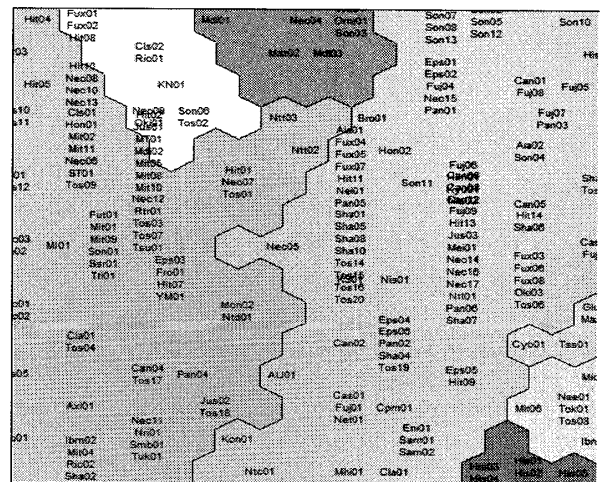


図 4 特許毎の特許マップ

5. おわりに

テキストマイニングと SOM を用いて、企業毎の特許マップと、特許毎の特許マップを作成した。その結果、視覚的に企業と技術の関係を調べることができた。今後は、特許毎のマップの分析、他分野のマップ作成、デンドログラム、k-means 法といった他のクラスタ分析や、2 軸マップとの比較を行なう予定である。

参考文献

- [1] 上田太一郎: 事例で学ぶ テキストマイニング, 共立出版(2008)
- [2] 酒井ほか: 特許明細書からの出願目的・技術課題情報の抽出, 人工知能学会全国大会(2009)
- [3] 平, 松本.: 特許調査支援システムの試作と評価, 人工知能学会全国大会(2009)
- [4] 西山ほか: 未来技術動向予測のための技術文書マイニング, 人工知能学会全国大会(2007)
- [5] T. Kohonen: Self-Organizing Maps, Springer-Verlag (1995)