
Appendix. 8 特徴語を用いた研究領域群の抽出

8-1 サイエンスマップにおける特徴語を用いた研究領域群の抽出とは

サイエンスマップにおいて、研究領域の内容を把握することは重要なステップである。そこで、Appendix 7 に記したように、論文のタイトルやアブストラクトを用いて各研究領域の特徴を示す語「特徴語」を抽出した。しかし、サイエンスマップ 2016 では研究領域数が 895 あり、それぞれの研究領域の特徴語に目を通すことは容易ではない。そこで、この特徴語を基に、ある程度同様の研究内容とみなせる研究領域群（複数の研究領域を包含したまとまり）を自動的に抽出することにより、サイエンスマップ全体の内容について把握できるように試みた。

なお、本調査で行った「特徴語を用いた研究領域群の抽出」のプログラム開発及びその運用については、VALUENEX 株式会社に委託し実施した。なお、特徴語の抽出方法に改良を加えたことに伴って、研究領域群の抽出に用いるパラメータ等についても調整を行っている。

8-2 サイエンスマップにおける特徴語を用いた研究領域群の自動選択アルゴリズム

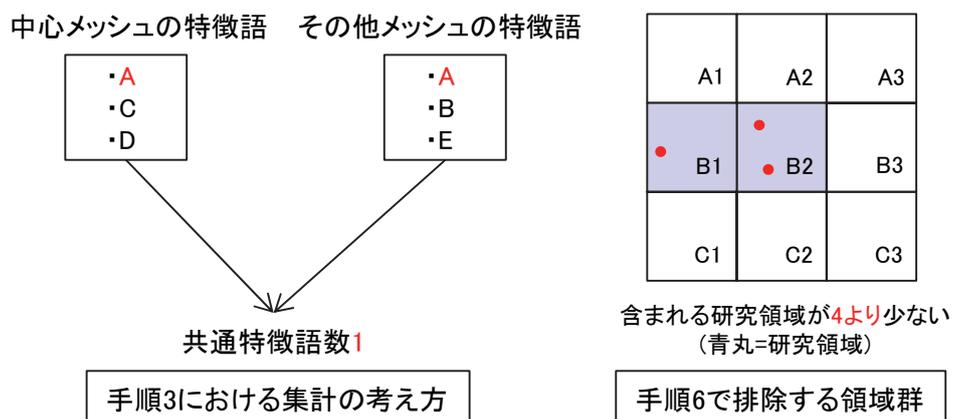
8-2-1 研究領域群候補の抽出

研究領域群候補の作成は、次に示す 7 つの手順によって行った。

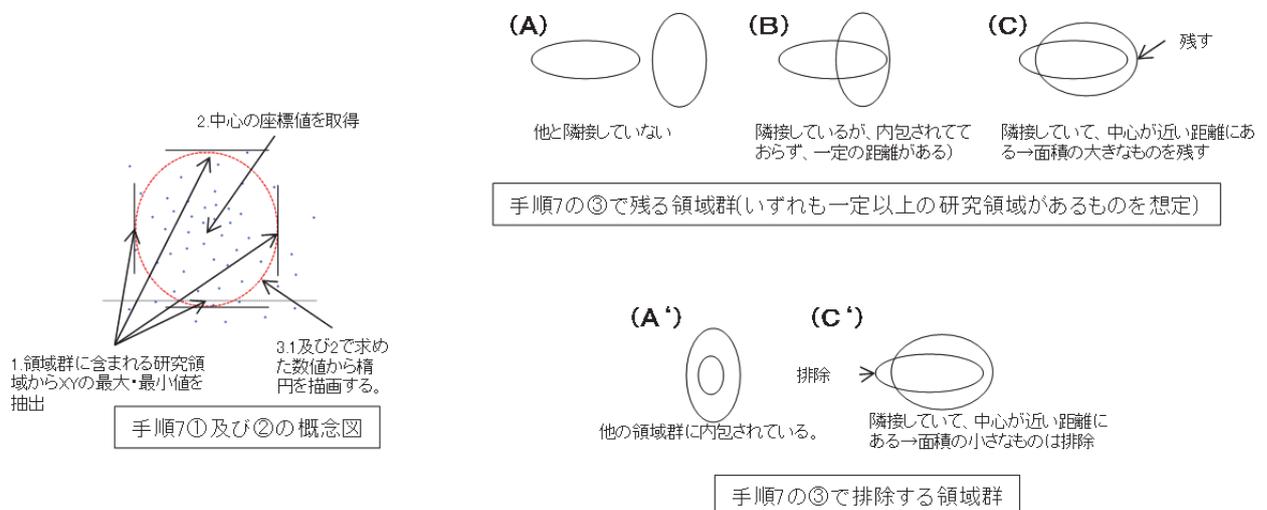
- (1) 手順 1: マップをメッシュ (400) に分割する。
- (2) 手順 2: メッシュに含まれる論文数 (密度) を計算する。
- (3) 手順 3: もっとも密度の高いメッシュについて、以下を行う。
 - ① 特徴語を集計する。
 - (ア) 「Appendix. 7」で抽出した「特徴語 (日本語訳)」を用いる。
 - (イ) 特定の特徴語が含まれる研究領域数を集計する。メッシュに 5 つの研究領域が含まれ、A という特徴語が 3 領域で出現する場合は 3、2 つの研究領域で出現する場合は 2 となる。比較に使用する特徴語の数は 60 を最大値とする。複数の領域が含まれ、特徴語の数が 60 以上あった場合、含まれる研究領域数が多い単語から上位 60 語が対象となる。
 - ② その他のメッシュに含まれる特徴語と比較し、同じ特徴語 (共通特徴語と呼ぶ) の件数をそれぞれ集計する。
 - ③ 共通特徴語が 1 以上であり、かつ、一定の範囲内 (距離 11, 注: 距離はメッシュ数) に含まれるメッシュを一つの領域群候補とする。
- (4) 手順 4: 手順 3 で領域群候補に設定されなかったメッシュの内、最も密度の高いメッシュについて、手順 3 と同様の処理を行う。
 - ① 手順 3 ですでに他の研究領域群候補に設定されたメッシュが選択されても、一つのメッシュが複数の領域群候補に属することを許すため、新たな領域群候補のメンバーに含める。
- (5) 手順 5: 手順 4 を実施すると、一定の距離範囲にあり、共通特徴語が 1 以上あるものは特定の研究領域群候補に属することとなる。いずれの研究領域群候補にも属さず、研究領域を含むメッシュが存在する場合、手順 4 を再実行する。研究領域を含み、いずれの研究領域群候補にも含まれないメッシュがなくなるまで手順 4 を繰り返す。したがって、研究領域群候補を作成する段階では、各研究領域はいずれかの研究領域群候補に所属する。

- (6) 手順6: 研究領域群候補に含まれる研究領域が4以下の研究領域群を削除する。
- (7) 手順7: 領域群候補について以下の処理を行う。
- ① 領域群に含まれる研究領域の中で、X、Y軸の最大値、最小値及び中心のXY座標を求める。
 - ② ①で求めた値について、中心のXY座標を中心とし、(X最大値-X最小値)をX方向の長さ、(Y最大値-Y最小値)をY方向の長さとした楕円を領域群の候補とする。
 - ③ 各領域群候補の候補に対し、仮定した楕円同士を比較し、以下の楕円を最終的な研究領域群として残す。
 - (A) 他の楕円に内包されない楕円である。
 - (B) 他の楕円と交差している楕円の内、中心点が一定以上離れている。
 ここでは、楕円の式が $X^2/A^2 + Y^2/B^2 = 1$ とした場合に、中心点 x_1, y_1 が $x_1^2/A^2 + y_1^2/B^2 > 0.5$ を対象とした。
 - (C) 他の楕円と交差し、楕円の中心点が一定距離以内にある場合、面積の大きな楕円を残す。

付録図表 8-1 手順3と手順6における考え方



付録図表 8-2 手順7の考え方



8-2-2 研究領域群候補の削除並びに統合による研究領域群の決定

ここまでのステップで得られた研究領域群のうち、類似のものを統合することで、最終的な研究領域群を決定した。