

研究組織における研究分野の可視化

竹野 凌¹⁾, 佐野 雅彦¹⁾, 松浦 健二¹⁾, 谷岡 広樹¹⁾, 大平 健司¹⁾, 関 陽介¹⁾, 上田 哲史¹⁾

1) 徳島大学

c501637027@tokushima-u.ac.jp

Visualization of Research Field in Research Unit

Ryo Takeno¹⁾, Masahiko Sano¹⁾, Kenji Matsuura¹⁾, Hiroki Tanioka¹⁾,
Kenji Ohira¹⁾, Yosuke Seki¹⁾, Tetsushi Ueta¹⁾

1) Tokushima University

概要

研究組織において、自組織及び学界での研究の動向を把握しておくことは重要である。本研究では、科学研究費助成事業による分野分類を用いて研究組織の研究分野を推定し、各分野における注力度合いを可視化する。

1 はじめに

研究組織の活動方針を決定する際に、自組織の注力している研究分野や学界で研究されている各分野の研究従事者数やテーマ数、研究内容（以下、研究動向）を把握しておくことは重要である。しかし、様々な研究が同時進行する中で把握を継続するには多くの労力を必要とする。

研究動向把握に関する先行研究は多く存在する。共引用分析を用いた研究分野の発展の時系列把握や研究者のポートフォリオ作成を行う研究[1]では光触媒分野においてその有用性が示されたが、他の分野や、複数分野における有用性は示されていない。また、分野に関わらず効果が期待されるテキストマイニングを用いた分析手法[2]では、入力されたキーワードを元に分析を行うため、分析対象となる研究分野に適切なキーワードを選択するための知識を必要とする。また、学界全体の中でこういった分野に位置するのか等は示されていない。

2 提案手法

研究組織において、全ての研究分野に関する知識がなくても自組織の研究がこういった分野に属するのか容易に把握できるよう、研究分野の推定及び可視化を行う。研究分野推定における先行研究として、科学研究費助成事業の分野分類を用いて CiNii に登録されている論文を分類する研究[3]があり、全ての学術領域を網羅している分類体系

として日本学術振興会が交付する科学研究費助成事業の研究分野分類が採用されている。

よって、本研究では、科学研究費助成事業による分野分類を指標として研究組織における各論文の研究分野の推定を行い、研究組織と科学研究費助成事業それぞれの研究動向を可視化することで研究組織において各研究分野への注力度合いの把握を容易にする。

2.1 研究分野推定

本研究では、分野分類の指標として、科学研究費助成事業による研究分野分類を用いる。具体的には科学研究費助成事業データベース(KAKEN)[4]から取得した、各分野における採択された研究題目に付与されているキーワードを集計したリストを各分野における特徴量とする。また、研究組織における論文タイトルとその抄録における、KAKEN の各分野のキーワードの出現回数によって研究内容が近い分野を推定し、論文の分類を行う。論文タイトル及び抄録には、論文の特徴が記載されており、出現する語句は分野判定に十分であると想定できる。

各分野における採択された研究題目のキーワードリストをそれぞれ1つの文書とする。出現単語とその出現回数を元に各分野の特徴量としてTF-IDF法による各単語への重み付け値を求める。重み付け値は式(1)で表される。キーワードリスト i の単語 j のTF-IDF値を w_{ij} とする。キーワードリスト i における単語 j の出現頻度を tf_{ij} 、キーワードリストの集合 D の総数を $|D|$ 、 D のうち単語 i

を含むものを df_i とする。

$$w_{i,j} = tf_{i,j} \log \frac{|D|}{df_i} \quad (1)$$

次に、論文のタイトル及び抄録に含まれる各分野のキーワードの $w_{i,j}$ の総和を計算し、各分野への関連度とする。関連度が高い分野ほど、その論文に研究内容が近いと想定される。また、論文ごとの抄録における単語数の影響を無くすために、論文1本ごとの全分野への関連度の合計が1となるよう正規化を行っている。

次に、論文ごとに求めた研究組織における全論文の関連度の各分野における総和を、研究組織におけるその分野の注力度とする。注力度が高いほど、その分野に関する論文が多く、研究従事者数やテーマ数が多いことが推定される。本研究の目的としては組織全体における各研究分野への注力度が把握できれば良いため、個々の論文の研究分野を推定するのではなく、全論文の関連度の各分野における総和を用いている。

2.1 可視化

学界全体の傾向を表す尺度として幅広く学術領域を網羅している科学研究費助成事業を用いる。前節で求めた研究組織における各分野への注力度と、科学研究費助成事業における各分野の採択課題数を可視化する。採択課題数を用いる理由としては、研究組織の現状と、採択により今後発表が期待される分野の傾向と比較するためである。スケールを合わせるため、どちらも百分率として計算した上で可視化している。また、D3.js を用い、全体を概観できるよう円環状に表示している。俯瞰して見ることで、こういった分野において科学研究費助成事業と比べ注力しているのか容易に把握することができる。

図1は可視化例として、2016年のKAKENにおける全321細目の採択課題のデータと徳島大学の論文データを元に注力度を可視化したものである。上が全体図、下が一部を拡大したものとなっている。円の中心からそれぞれ系・分野・分科・細目の分類となっており、円の外側にある棒グラフが注力度を表す。緑色がKAKENの採択課題数、赤色が徳島大学における注力度である。データとしては、KAKENから取得した2016年に採択された採択課題31110件、徳島大学のEducation and Research Database (EDB)[5]から取得した2016年の日本語論文2620本のデータを用いている。

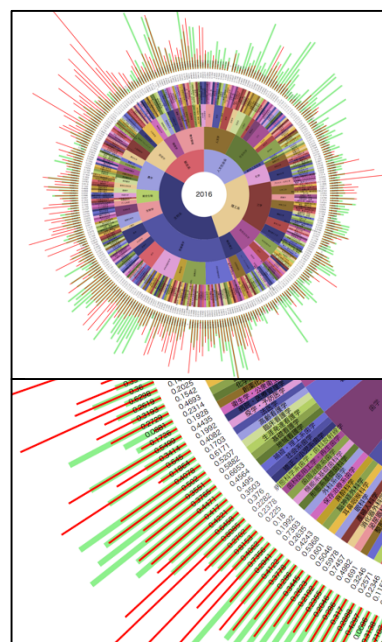


図1. 可視化例 (2016年)

3 おわりに

本研究では、研究組織の研究動向把握を容易にするため、科学研究費助成事業における研究分野分類を用い、研究組織における論文の研究分野の自動推定を行い、研究組織と科学研究費助成事業の各研究分野への注力度の可視化を行った。今後の展望として、各分野内での頻出キーワードや研究内容の移り変わり等を分析する機能を追加することで、研究動向把握をさらに容易にすることが可能であると考えられる。

謝辞

本研究はJSPS 科研費 JP15K12168 の助成を受けたものです。

参考文献

- [1] 七丈 直弘, “共引用クラスタリングによる研究分野の動的把握に向けた試論”, 情報知識学会誌, Vol.23, No.3, 2013.
- [2] 増田 勝也, 丹治 信, 植松 すみれ, 美馬 秀樹, “研究動向分析のための論文のデジタルテキスト化とマイニングシステム”, 言語処理学会第20回年次大会発表論文集, 2014.
- [3] 福田 悟志, 難波 英嗣, 竹澤 寿幸, “要素技術とその効果を用いた学術論文の自動分類”, DEIM Forum, 2015.
- [4] “KAKEN 科学研究費助成事業データベース”, <https://kaken.nii.ac.jp/ja/index/> (参照 2017-9-27)
- [5] “徳島大学 教育・研究者情報データベース”, <http://web.db.tokushima-u.ac.jp> (参照 2017-9-27)